

AD-A217 332

Technical Report
860Radiation-Hardened
Wafer Scale IntegrationA.H. Anderson
J.A. Burns
C.K. Chen
G.J. Dunn
K.H. Konkle
B-Y. Tsaur
C.E. Woodward
P.W. Wyatt

25 October 1989

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS



Prepared for the Department of the Air Force
under Contract F19628-90-C-0002.

Approved for public release; distribution is unlimited.

DTIC
ELECTE
JAN 30 1990
S E D

90 01 30 094

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

RADIATION-HARDENED WAFER-SCALE INTEGRATION

*A.H. ANDERSON
J.A. BURNS
G.J. DUNN
K.H. KONKLE
C.E. WOODWARD
P.W. WYATT
Group 23*

*C.K. CHEN
B-Y. TSAUR
Group 83*

TECHNICAL REPORT 860

25 OCTOBER 1989

Approved for public release; distribution is unlimited.

LEXINGTON

MASSACHUSETTS

EXECUTIVE SUMMARY

A focal plane processor (FPP) for a large array of LWIR photodetectors on a space platform must process large amounts of data, operate reliably in a high-radiation environment, occupy small space, and use little power. This project had the goal of demonstrating that these objectives could be achieved with wafer-scale (WS) circuit integration on silicon-on-insulator (SOI) wafers. The WS integration technology is Lincoln Laboratory's Restructurable VLSI which uses a laser to form connections and make cuts on two levels of metal. Wafers are fabricated with unconnected circuits and WS interconnect, and after testing the laser is used to alter circuits and to wire together good circuits to achieve customization and defect avoidance. The technology and design tools have been demonstrated through development of six different wafer-scale systems. A technology [called Zone-Melting Recrystallization (ZMR)] for making oxide-isolated wafers had been developed at Lincoln Laboratory and circuits have been fabricated in these wafers. There were four elements in this program: (1) to design a prototype WS FPP, (2) to improve the ZMR process, (3) to develop a CMOS fabrication process in either ZMR or SIMOX SOI wafers, and (4) to fabricate and restructure the WS FPP. The first three elements were accomplished, but the program was terminated before the wafer-scale circuit was fabricated.

This prototype system was designed to handle a 5-column, 64-row scanning detector array in which each detector is sampled every $7\ \mu\text{s}$. Because of the relatively low data rate and to minimize wafer-scale interconnect, a serial architecture was used and the system was partitioned into 8 identical processors. To allow fault tolerance 10 processors are provided; an external controller can test them and set multiplexors so that any 8 of the 10 can be used. The processors are hardwired to (1) perform a unique 4-segment offset and gain correction for each detector, (2) delay signals from the 5 columns to time align signals from 1 target, (3) recognize and reject signals which may be γ -corrupted, (4) average 'good' signals, and (5) perform a 4×4 filter function and threshold the result. The data correction coefficients, filter kernel, threshold, and γ constant are loaded through a serial bus. Laser restructuring is used to give each circuit on the wafer a unique bus address. Since the circuits were to be built in an experimental process, each replaceable circuit was limited to less than 12,000 transistors which resulted in 5 different circuits. Static CMOS circuitry was used for radiation resistance. All 5 circuits were designed and built in a $3\text{-}\mu\text{m}$ bulk process through the MOSIS silicon foundry. The circuits operated above the design clock rate of 16 MHz, and yields were very high in this mature process. With $2 \times$ circuit redundancy for 4 circuits and $1.6 \times$ for the smallest circuit, 5 processors can be fit onto a $45\text{-} \times 41\text{-mm}^2$ area on a 3-in wafer so that 2 wafers would be required for the 10-processor system. Each wafer would be packaged in a 2-in square package which has been used for earlier wafer-scale circuits. On a 5-in silicon wafer with $2\text{-}\mu\text{m}$ processing, enough cells could be built to place 40 or 50 processors on a wafer.

The starting material for the ZMR process is a silicon wafer with a thermally grown SiO_2 film, a poly-Si layer formed by low-pressure chemical vapor deposition and a capping layer of SiO_2 . The entire wafer is heated to a base temperature below the melting point of Si and a movable strip heater is used to produce a narrow molten zone in the poly-Si layer. As the zone is translated, a recrystallized Si film is formed. The original graphite strip heater system for the ZMR process

produced device-quality material but did not yield uniform films routinely, could not produce films on wafers larger than 3 in. and the films contained a high density of subboundaries. A new ZMR system was built which provides a uniform base temperature over the entire Si film surface, has a constant spacing between the upper strip heater and the wafer throughout the heater scan, and has improved mechanical, chemical, and thermal stability. These improvements resulted in a smooth motion of the liquid-solid interface over a wide range of scan speed and improved run-to-run reproducibility. Four-inch wafers have been recrystallized to within 3 mm of the perimeter, free of edge-related macroscopic defects. An earlier study showed that a SiN_x film on top of the SiO_2 cap promoted wetting of the molten Si zone. An improved process in which the SiO_2 cap is annealed with NH_3 results in a more uniform and better controlled N concentration at the SiO_2 interface, and with this technique, films have been produced that are extremely smooth and uniform in thickness. These improvements result in films with few, if any, subboundaries. Promising results were obtained with a new ZMR configuration which has the stationary and movable heaters on the same side of the wafer.

An existing integrated circuit fabrication process for 2- μm CMOS in bulk silicon was adapted for 3- μm CMOS on SOI wafers with a goal of transient radiation hardness consistent with SDI Level I goals and total dose hardness of 1 Mrad(Si). Transistors are isolated by etching nontransistor material down to the buried oxide. Satisfactory coverage of the poly-Si gate over the steep, 300-nm-high island edges was achieved. N-channel sidewall threshold was increased by implanting boron into field areas adjacent to n-channel transistors and diffusing it laterally into the transistors before doing the mesa etch. Back channel hardness has been demonstrated but further work is needed on side-channel hardness. Gate oxide breakdown voltage was low on early ZMR films, but for newer films the breakdown voltage is 80 to 90 percent that for an oxide on bulk silicon. Subboundaries in the older ZMR material appear to cause leakage paths between source and drain, but this leakage has not been observed in the newer, subboundary-free material. Wafer length interconnect and vertical links have been successfully made on SOI wafers. Fabrication in SOI is continuing for two of the FPP circuits; results will be reported in a project memorandum.

A separately funded project to develop an extremely hard gate dielectric by the nitridation of SiO_2 is reported because of its relevance to this program. The dielectric is produced by first growing a conventional oxide of the desired thickness and then, in the same furnace tube, partially converting it to a nitride by exposure to ammonia (nitridation), followed by a second oxidation. The process can easily be incorporated in a typical fabrication sequence. We have demonstrated a 37-nm dielectric which exhibits zero interface state increase and only -1.35-V threshold voltage shift after 100 Mrad(Si), very high resistance to channel hot carrier stress, and a factor of seven improvement in charge-to-breakdown (Q_{bd}) over conventional oxide.

The focal plane array postulated by the sponsor of this project comprised 400,000 detectors with a data processing requirement of 10^{12} operations per second, with radiation tolerance of SDI Level II, and a level of fault tolerance consistent with operation in orbit for many years. To accomplish that computational throughput, a special-purpose processor was proposed in order to avoid the $\sim 10\times$ penalty in size, weight, and power typical of general-purpose processors. Extrapolation from existing wafer-scale devices indicates that such a system would occupy ~ 25 6-in wafers built

with 1- μ m technology. A system built with conventional ICs and packaging would be very much larger. The required radiation dose rate and single-event upset tolerance of SDI Level II are very difficult to accomplish in bulk silicon, but are relatively straightforward in SOI. Reoxidized nitrided oxide can readily meet the total dose requirement of Level II. These considerations led to the approach which is described in this report. The prototype processor designed in this project would have demonstrated *all* the postulated capabilities, including fault tolerance, Level I radiation hardness, and inter-wafer communication, but for a small array of detectors. Because of the modularity of the design, extension to a larger array would entail nothing more than adding more wafers.

A difficulty with any special-purpose processor is that if the system requirements change, then the processor must be redesigned. This report describes a highly modular architecture in which the modules are relatively simple. A new system definition would require changing the number of those modules and might require redesign of some of them, but since the modules are small, the effort would be correspondingly small. It seems certain that large, scanning LWIR arrays will once again be of interest in the future, though their specifications will differ from those postulated for this study. The architecture and technology discussed here should be readily adaptable to those new requirements.



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

TABLE OF CONTENTS

EXECUTIVE SUMMARY	iii
LIST OF ILLUSTRATIONS	ix
LIST OF TABLES	xi
1. INTRODUCTION	1
2. SYSTEM DESIGN	3
2.1 Function	3
2.2 Architecture	3
2.3 Partitioning	3
3. CIRCUITS	7
3.1 General	7
3.2 Input and Calibration	11
3.3 Delay	18
3.4 Threshold	20
3.5 Average	26
3.6 Filter	31
4. WAFER DESIGN	43
4.1 Floor Plan	43
4.2 Routing Experiments	44
5. SOI Films by ZMR	47
5.1 ZMR by the Graphite-Strip-Heater Technique	47
5.2 New System for ZMR	48
5.3 New Capping Technique for ZMR	52
5.4 Liquid-Solid Interface Morphologies During ZMR	56

5.5	ZMR With Enhanced Radiative Heating	62
5.6	Technology Transfer and Remaining Materials Issues	65
6.	DEVELOPMENT OF A WAFER SCALE IC PROCESS FOR SOI SUBSTRATES	67
6.1	Background	67
6.2	Technical Issues	67
6.3	Experimental Results	68
6.4	Summary and Outstanding Issues	73
7.	RADIATION-HARDENED REOXIDIZED NITRIDED OXIDES	75
8.	CONCLUSIONS	81
	REFERENCES	83
	APPENDIX - Design of a Wafer-Scale Focal Plane Processer	85

LIST OF ILLUSTRATIONS

Figure No.		Page
2-1	Schematic Wafer Layout	4
2-2	Block Diagram of One Processor	6
3-1	Basic Circuit Elements	8
3-2	Clocked Circuit Elements	9
3-3	Delay Between Circuit Cells	10
3-4	<i>Input</i> Block Diagram	12
3-5	<i>Input</i> Circuit Implementation	14
3-6	Circuits in the Multiplier-Adder	16
3-7	Multiplier and Adder Circuits	17
3-8	<i>Delay</i> Block Diagram	18
3-9	<i>Delay</i> Circuit Implementation	19
3-10	<i>Threshold</i> Block Diagram	21
3-11	<i>Threshold</i> Circuit Implementation	22
3-12	Five-Input Adder Circuit	24
3-13	Parallel Description of Two Adder Stages	25
3-14	<i>Average</i> Block Diagram	27
3-15	<i>Average</i> Circuit Implementation	28
3-16	<i>Average</i> Coefficient Generator	29
3-17	<i>Filter</i> Block Diagram	32
3-18	<i>Filter</i> Circuit Implementation	34
3-19	Quad Multiplier-Adder Circuit	35
3-20	Multiply-and-Round Stage	36
3-21	Combinatorial Full Adder	36
3-22	Add-and-Round Stage	37
3-23	Synchronization Circuit	38

Figure No.		Page
4-1	Wiring Diagram of One FPP Wafer	43
4-2	Floor Plan of the Top Half of the FPP Wafer	45
4-3	Routing of Signals on an FPP Wafer	46
5-1	Schematic Diagram of Graphite-Strip Heater	48
5-2	External View of New System for ZMR of SOI Films	49
5-3	Top View of Interior of New ZMR System	49
5-4	Schematic of Computer Control of the Lower and Movable Upper Heaters	51
5-5	Schematic Cross Section of a Typical SOI Wafer	52
5-6	Optical Micrographs of a Recrystallized SOI Film with Monochromatic and White Light	53
5-7	Optical and Electron Micrographs of SOI Film	54
5-8	Optical Micrograph at Higher Magnification	55
5-9	Interface Morphologies as a Function of Upper-Strip Power	56
5-10	Cell Spacing for Four Grains as a Function of Upper-Strip Power	57
5-11	Optical Micrographs Showing Three Types of Defects	58
5-12	Fraction of Three Defect Types Versus Power	59
5-13	Interface morphologies as a function of increasing velocity	60
5-14	Dependence of Cell Period on Zone Velocity	60
5-15	Optical Micrographs of Defect Patterns	61
5-16	Cross-Sectional Diagrams of Conventional and New ZMR Configuration	63
5-17	Optical and Electron Micrographs of Subboundary-Free Defect-Etched SOI Film	63
6-1	Drain Current Versus Gate Voltage for Two n-Channel ZMR Transistors	69
6-2	Simulated Doping Profile of a p-Channel Transistor Body	71
6-3	Subthreshold Leakage Current Believed to be Associated with Subboundaries in Older ZMR Material	72
7-1	Midgap Interface State Density Versus X-Ray Dose in Capacitors	76
7-2	Midgap Voltage Shift Versus Applied Oxide Field in Irradiated Capacitors	77

Figure No.		Page
7-3	Midgap Voltage Shift Versus X-Ray Dose for Capacitors	78
7-4	Midgap Voltage Shift Versus X-Ray Dose in p-Channel Transistors	79
7-5	Normalized Inversion Layer Mobilities of p-Channel Transistors	79
7-6	Transconductance Degradation	80

LIST OF TABLES

Table No.		Page
2-1	FPP Circuit Cells	5
3-1	Coefficient for Different Number of Good Inputs	40
3-2	Filter Output Timing	41

1. INTRODUCTION

The prototype wafer-scale (WS) focal plane processor (FPP) is a demonstration of the application of the Lincoln Laboratory Restructurable VLSI technology [1] to the requirements for massive parallelism in the initial signal processing of data from an array of photodetectors. It was also to be the first implementation of a WS circuit in a radiation-hard semiconductor process. The project was not continued to fabrication of the WS circuit due to funding limitations and changes in priorities. This report describes the design of the circuits and presents test results from their fabrication in a bulk CMOS process. The preliminary plan for the WS design is presented.

Restructurable VLSI comprises a methodology, technology, and a set of CAD tools for building large area integrated circuits (ICs). Wafers are fabricated with redundant circuits and interconnect. both circuits and interconnect are tested after fabrication, and a laser is used to connect the operable circuits to build the desired system. The laser can also be used to customize circuitry, for instance to set the coefficients in cells used to implement a filter function. Several laser restructuring technologies have been developed. One technique which uses a laser to form a connection between two adjacent diffusions is completely compatible with standard IC processing. Another approach forms connections between two layers of metal and uses a silicon nitride film between the metal layers. Laser-created connections and metal cuts are made with high yield and appear to be very reliable. Accelerated aging tests have been done with favorable results and one WS circuit has operated without failure in a laboratory for more than four years. WS systems have been built which are 50 mm on a side and contain 400,000 active transistors. In design at Lincoln Laboratory are larger systems, which will have 3,000,000 active transistors on an 80-mm² piece of silicon.

New techniques and equipment have been developed in this program for the preparation of SOI films by zone-melting recrystallization (ZMR). CMOS circuits are being fabricated in this material and we intend to complete fabrication of the filter and delay circuits. Radiation-hard circuits also require special gate dielectrics and a summary is given of results from a related research program on reoxidized nitrided oxides.

2. SYSTEM DESIGN

2.1 FUNCTION

The FPP is an array of processors, each of which performs operations on signals from a 5×8 array of photodetectors. A previous report [2] (attached as an Appendix) describes the system for which the processor was designed and some of the issues of system partitioning. This prototype system has 8 active and 2 spare processors. An off-wafer controller substitutes processors by setting switches in the wafer circuitry. The processors are largely independent except for data connections required for spare substitution and for the filter function. With $3\text{-}\mu\text{m}$ technology, 5 processors can be placed on a 3-in wafer.

Figure 2-1 is a block diagram of 1 wafer. Each processor comprises one column in this figure and each processor has inputs from 5 columns of detectors. These are the 5 groups of A, B, C, D, E inputs. The square blocks perform switching of inputs for spare substitution, signal correction for detector offset and nonlinearity, and a time delay which is different in each row (represented by TDI0-4). At each sample period the 5 inputs presented to the circuits, represented as circles, are signals from 5 detectors for the same target point. This circuit rejects signals which are statistically too big, signals which are probably corrupted by gamma rays (see [2] for details), and produces an average of the remaining signals. The triangle circuit performs a 4×4 spatial filter function for which purpose signals from adjacent processors are required. The diamond function switches the outputs of the 4 currently active processors to output pins and was not implemented in this design.

2.2 ARCHITECTURE

In the postulated system data from eight detector rows are to be processed in $7\text{ }\mu\text{s}$ or 875 ns per detector. Input data are assumed to be 12 bits long. A serial architecture was chosen for these reasons: (1) with parallel data buses the interconnect shown on Figure 2-1 would take up a large amount of wafer space, (2) the partitioning issues discussed in Section 2.3 favor serial architecture, and (3) the relatively low data rates make inefficient use of a parallel architecture. To accommodate overflow in the arithmetic circuitry, an internal 14-bit word is used; it is assumed that input data are presented to the wafer in 14-bit serial form with the the least significant bit (LSB) first and the 13th and 14th bits set to ZERO. All inputs are in phase and a synchronization signal is present which has a ONE in the LSB time slot of each word. In the first word of each 8-word frame of data, the synchronization signal is ONE for 2 bit times starting with the LSB. The processors create an output synchronization signal.

2.3 PARTITIONING

System partitioning is a critical issue in RVLSI design. If the circuit cells are too large for a given fabrication technology, then their yield will be low and many extra cells must be built on the

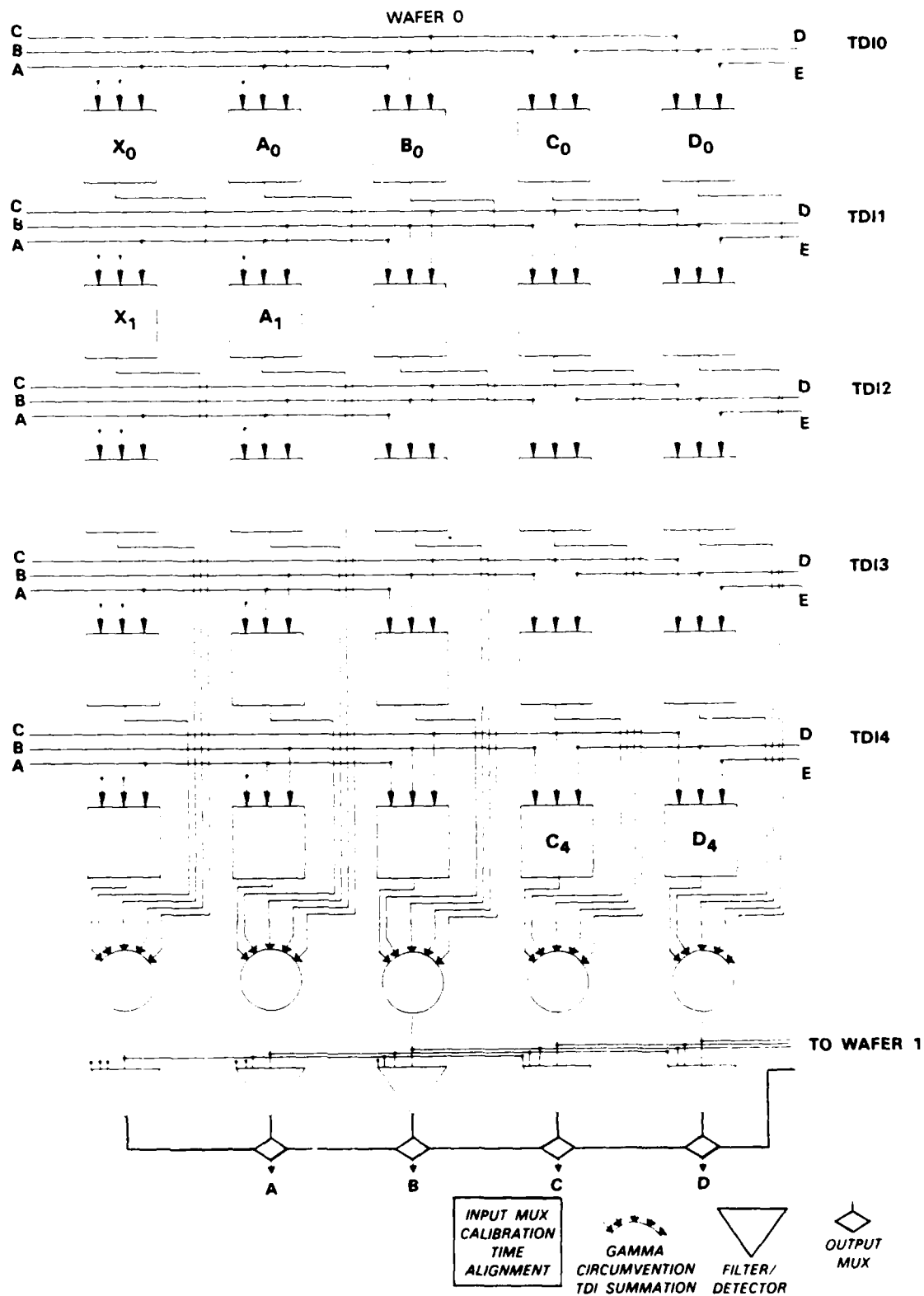


Figure 2-1. Schematic wafer layout.

wafer for each used circuit, leading to inefficient use of wafer area. On the other hand, if the cells are too small then a large amount of WS interconnect will be needed, again leading to inefficiency. Other current designs, which are to be fabricated in a mature 2- μ m CMOS technology, have circuit cells with 50,000 transistors in logic circuitry and 100,000 transistors for memory circuits. A cell size which results in about 50 percent yield seems to be about right [3]. Since the FPP wafer was to be fabricated in an experimental SOI technology, it was decided to limit cell size to about 12,000 transistors. This limit was rather arbitrary, it may be ambitious from a fabrication standpoint, but it was quite satisfactory for partitioning. When the testing results are presented, it will be seen that for CMOS bulk fabrication the cells are smaller than necessary. With this constraint the system was partitioned into five cell types: *input* and *delay (tdu)*, for the square cell function of Figure 2-1, *threshold* and *average* for the circle function, and *filter* for the triangle function. Each row of square cells in Figure 2-1 must have a different delay; TDI0 has 4×32 word delays, TDI1 3×32 , TDI2 2×32 , TDI3 1×32 , and TDI4 0. The *delay* unit was built to have 32 words of delay and units are cascaded to achieve longer delays. Table 2-1 shows the size of each cell and the number required for each processor. The block diagram of one processor is shown in Figure 2-2.

TABLE 2-1.
FPP Circuit Cells

FPP CELLS			
Cell	No/Proc	Size (mm)	Transistors
<i>Input</i>	5	2.5 \times 3.6	11100
<i>Delay</i>	10	1.2 \times 2.9	4200
<i>Threshold</i>	1	2.5 \times 3.0	6700
<i>Average</i>	1	2.4 \times 4.0	10400
<i>Filter</i>	1	2.5 \times 4.0	11800

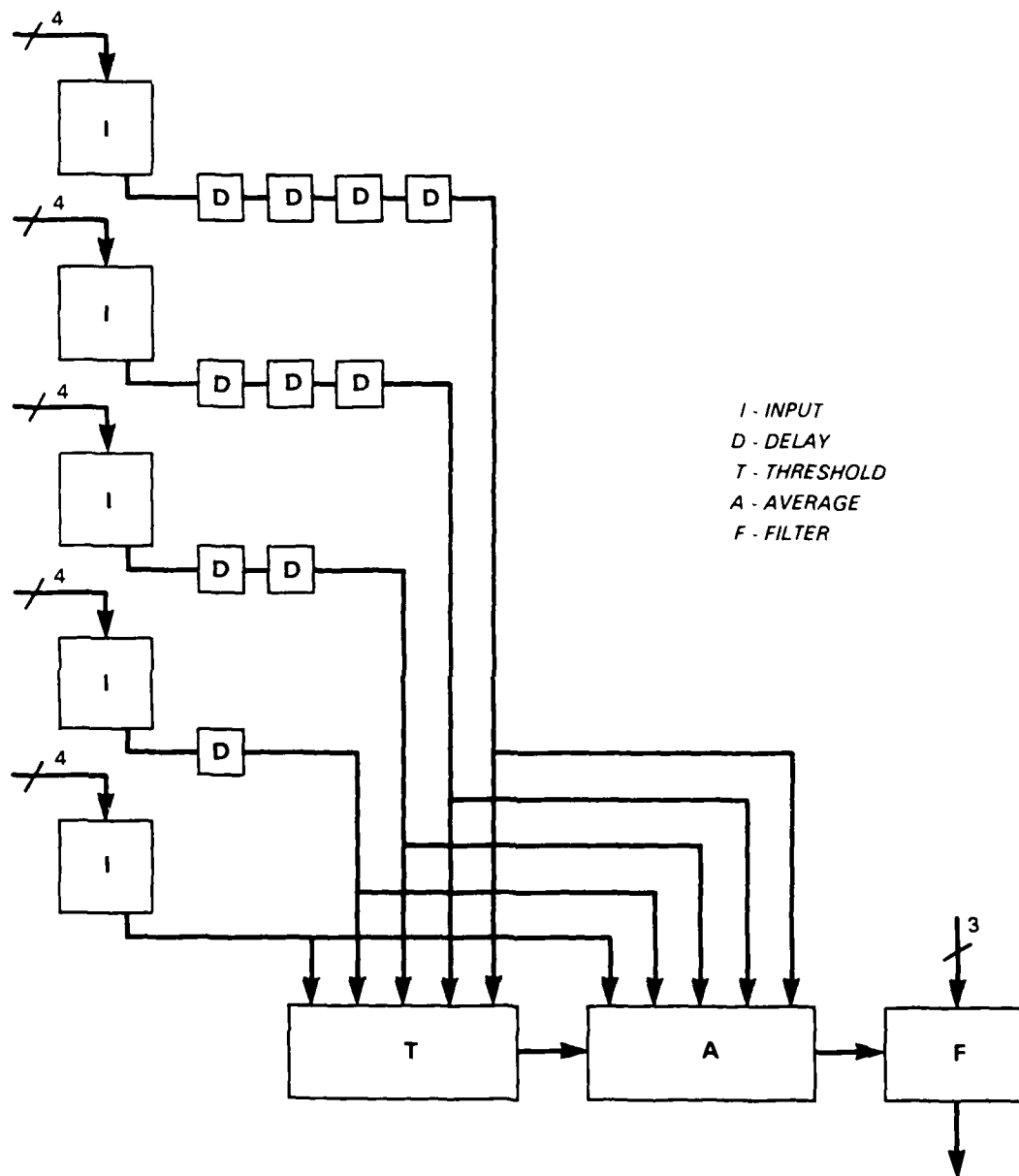


Figure 2-2. Block diagram of one processor.

3. CIRCUITS

3.1 GENERAL

3.1.1 Technology and Environment Considerations

The FPP is designed to be fabricated in any of four technologies. For reasons of radiation hardening, SOI is primary and uses wafers prepared either by ZMR or separation by implanted oxygen (SIMOX)[4]. Fabrication can also be in either P-well or N-well bulk silicon. The experimental results presented in this report are from packaged circuits fabricated in $3\mu\text{m}$ P-well CMOS through the MOSIS foundry service. All data storage in the circuits is static for compatibility with a radiation-resistant design philosophy.

3.1.2 Design and Testing

Circuit layout was done with the Magic program [5] on Sun Microsystem, Inc. workstations. The COSMOS program [6] was used to do logic simulations on circuits extracted from the layouts. Some simulation of arithmetic behavior was done with special C-language programs. Preliminary wafer floor planning was done using the SLASH programs [7] developed in the Lincoln Laboratory RVLSI program.

Packaged circuits were tested on a Tektronix S3260 tester. The same tester and programs would be used with probe cards to test the circuits on a wafer.

3.1.3 Circuit Elements

Standard logic circuits are implemented as combinational CMOS circuits, such as the inverter, NAND, and NOR circuits shown in Figure 3-1. Pass transistor logic is used in the transfer gate and XOR circuit as shown. The 20-transistor, master-slave flip-flop shown in Figure 3-2(a), is a conservative design appropriate for a circuit which will be fabricated in any of several technologies and used in a radiation environment. Clocked-inverter transistors in the flip-flop are $4.5\mu\text{m}$ wide and $6\mu\text{m}$ in the inverters. A buffered flip-flop, Figure 3-2(b), has wider transistors in the feed-forward inverters of the slave latch. Two different implementations of a selector are used. One version, Figure 3-2(c), has two transmission gates at the input of an FF; the second, Figure 3-2(d), incorporates a clocked AND-OR gate in place of the input clocked inverter of the master latch. The acceptor circuit of Figure 3-2(e) is a latch clocked the same as an FF slave circuit. Its use is described next.

3.1.4 Clocking

All circuits are clocked from a clock signal distributed on the wafer and buffered in each cell. The internal circuits store all data in master-slave flip-flops designed to accept a new datum when

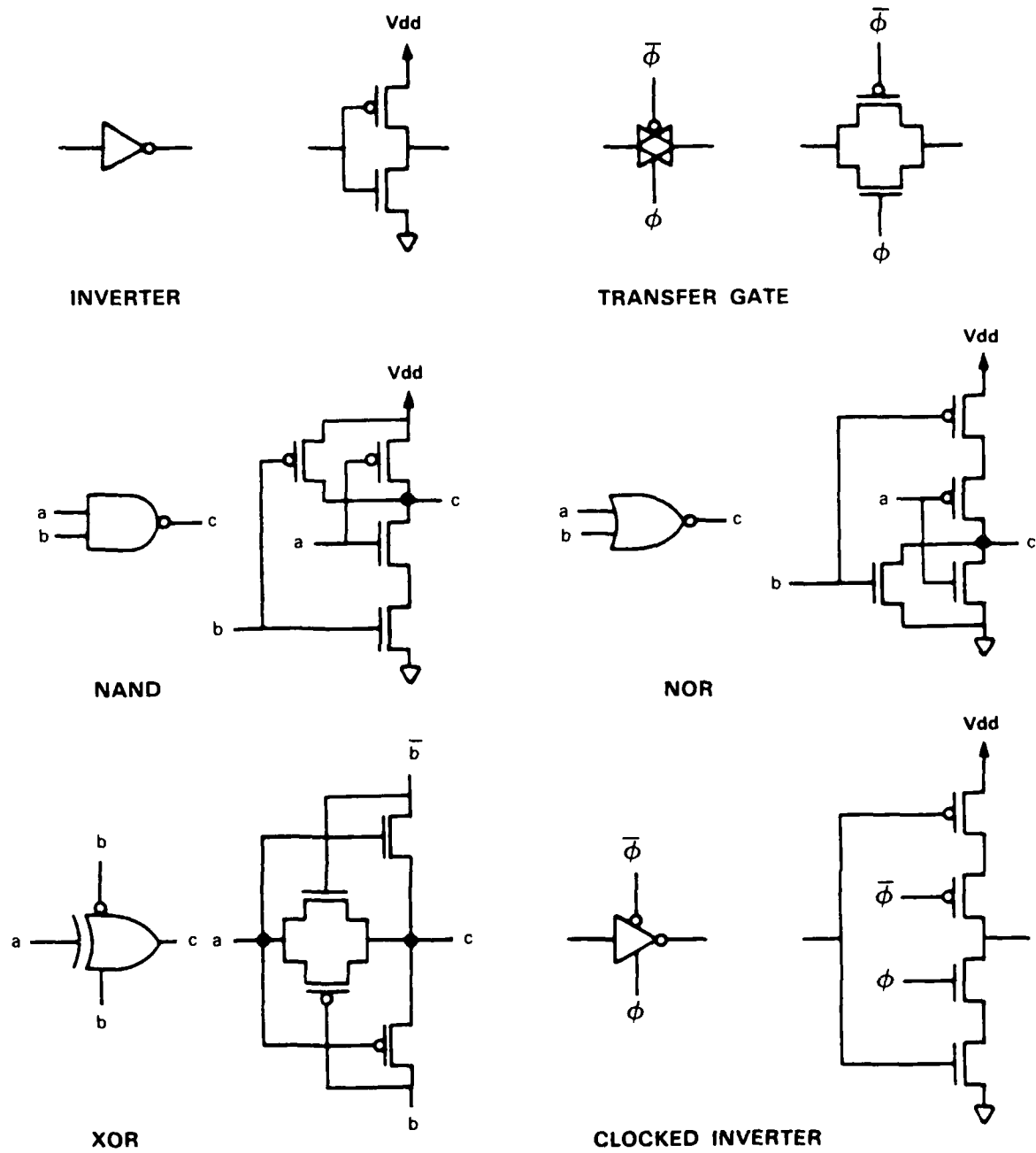
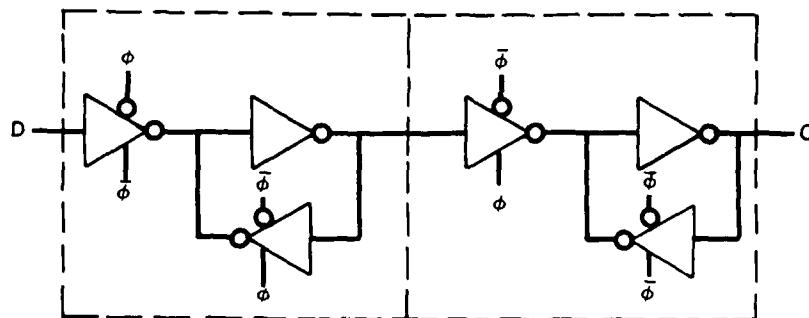
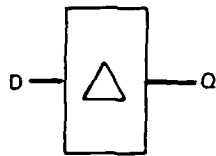
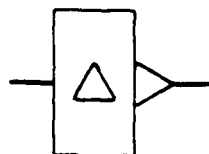


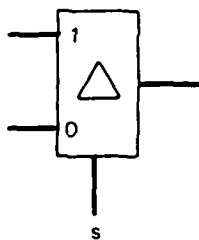
Figure 3-1. Basic circuit elements.



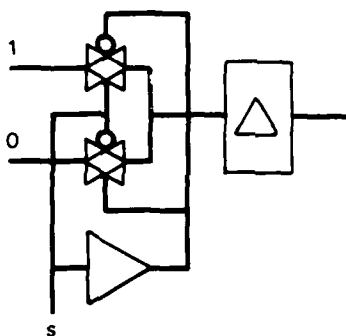
(a) FLIP-FLOP



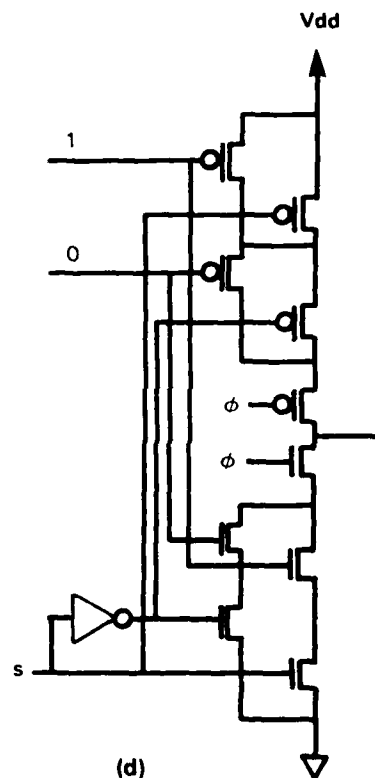
(b) BUFFERED FLIP-FLOP



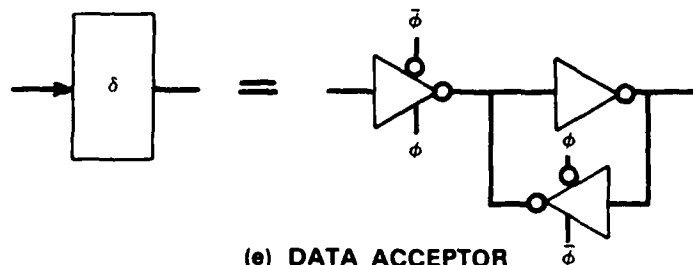
SELECTOR FF



(c)



(d)



(e) DATA ACCEPTOR

Figure 3-2. Clocked circuit elements.

the clock is low, and latch and present it at the output when the clock goes high; data sampling and output both occur on the rising edge of the clock. This has certain implications for the timing of external signals. Input circuits of the cells are designed to accept data under conditions of clock skew that might occur on a WS circuit.

In Figure 3-3(a), the connection from the output of one FF to the input of another represents a data path between circuit cells A and B on a wafer. The sum of the delays through an output

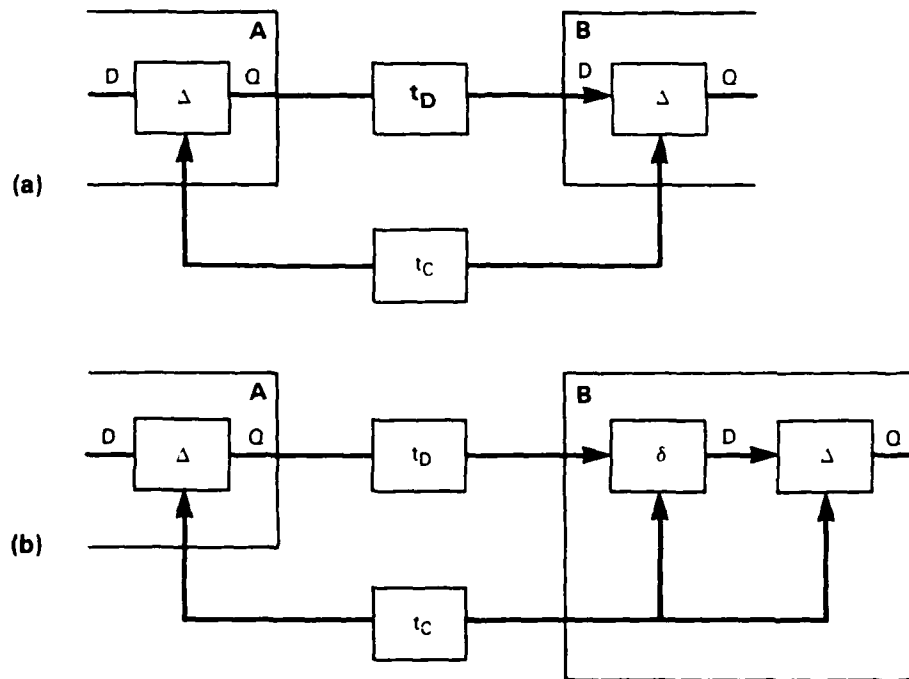


Figure 3-3. Delay between circuit cells.

buffer, wiring, input buffer, and any combinational logic in the path is represented by t_D , which is always positive. Delay t_C represents the relative timing of the clock edge in cell B to that in cell A and is a measure of skew in clock distribution on the wafer and differences in delay in clock buffers and wiring in the two cells; it may be positive or negative, ideally it is zero. Let the required setup time for a latch be t_s and the clock period be t_P . Then, in Figure 3-3(a):

$$-(t_P - t_D) < t_C < t_D - t_s.$$

If t_D is small compared to t_P , then positive clock skew t_C , is very limited. To improve this situation the acceptor circuit was inserted in the input path as shown in Figure 3-3(b). This circuit is a slave latch from Figure 3-2(a). If the clock signal has a 50-percent duty cycle, then the limits on t_C are:

$$-(t_P/2 - t_D - t_s) < t_C < t_P/2 - t_D.$$

If t_P is large relative to other delays, this gives more margin on the positive side. Whether the acceptor circuit is advantageous will be determined by the values of t_D and t_C . A decision on whether to use the acceptor circuit would have been made after all circuit measurements had been done and a WS layout had been completed.

3.1.5 Parameter Setting

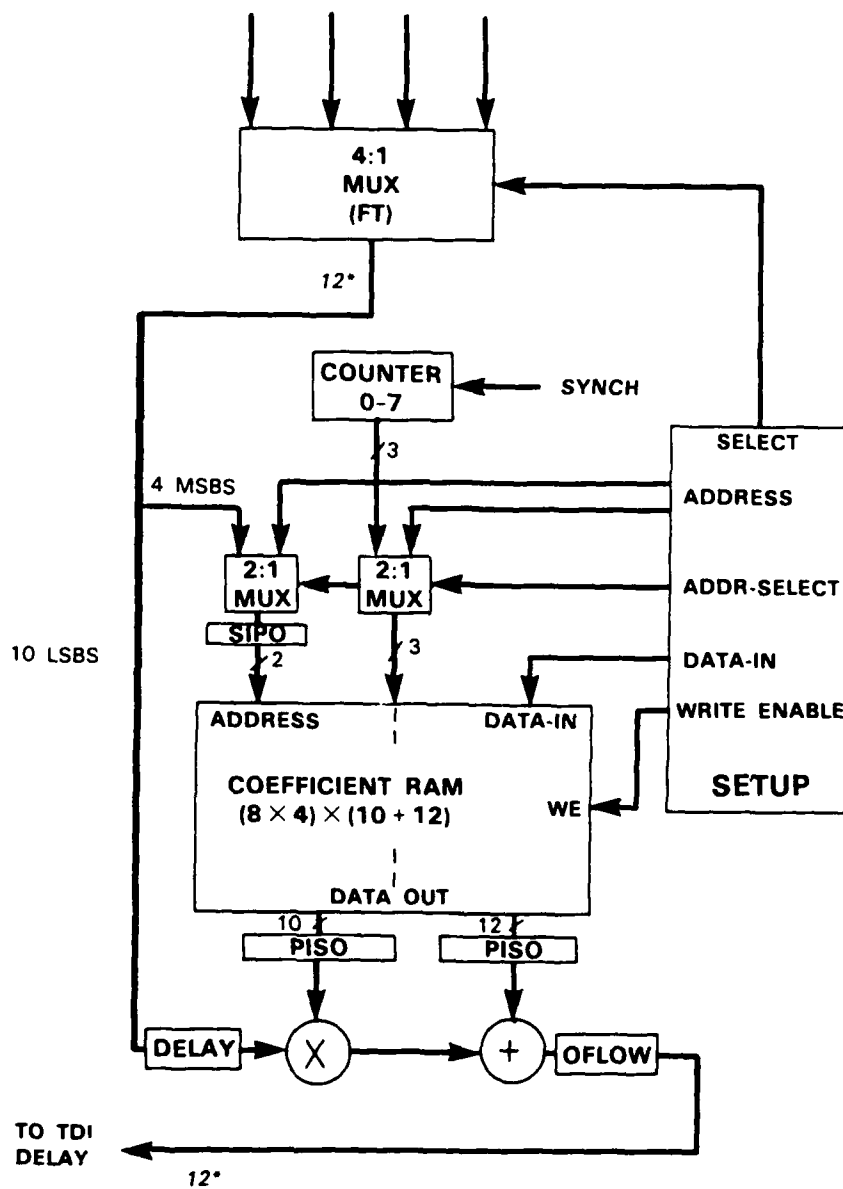
Three of the five types of cells have parameters which are set during operation by an off-wafer controller. The *input* cell requires 704 bits of data for the coefficient memory and two bits to select a data input line on the 8-of-10 sparing. *Threshold* has a 5-bit gamma coefficient and *filter* also has a 2-bit selector control, 16 bits for the kernel of the 4×4 filter function, and 14 bits of threshold data. Parameters are set when the processors are not doing data operations and they can be done slowly.

A 3-line broadcast bus is used to serially load parameters. The 3 lines are the clock **iclk** (which is independent of the clock for data circuits), a control line **imode**, and a data line **data**. This bus is connected to each of the active *input*, *threshold*, and *filter* cells. Each cell is given a unique 7-bit address by laser-formed links that set the referent word. For test purposes the referent is set by levels applied to probe pads. A bit-serial address is applied to **data** with **imode** TRUE. When **imode** is made FALSE, the cell address recognizer compares the 7 most recent bits with the referent. If they match, it will accept data until **imode** is again TRUE.

3.2 INPUT and CALIBRATION

3.2.1 Function

Figure 3-4 is a block diagram of the *input* circuit. Each processor has 5 of these circuits, 1 for each column of the detector array. The input MUX is set by an off-wafer controller and selects 1 of 3 inputs, dependent on which 8 of the 10 processors are being used or a 4th input which may be a source of test signals. The input is bit serial, 14 bits, LSB first with the 2 MSBs always at ZERO. The principal function of this circuit is to correct the data for detector nonlinearity and offset. There is a separate set of calibration coefficients for each of the 8 detectors assigned to an *input* circuit, and each calibration function has 4 linear segments. For each input datum, the appropriate slope and offset are selected by addressing the coefficient memory with an 8-count counter concatenated with the 11th and 12th bits (two most significant nonzero bits). The slope coefficients are stored with 10-bit accuracy which is sufficient to maintain input accuracy, since each coefficient is applied over 1/4 full range. The 10-LSBs of the input are multiplied by the 10-bit slope coefficient and added to the 12-bit offset coefficient. An overflow circuit sets outputs larger than 4095 to this maximum value [2]. The coefficient memory is loaded through the parameter setup bus.



* PROCESSOR IS SERIAL WITH BASIC PERIOD OF 14 BITS. THIS NOTATION REFERS TO NUMBER OF NONZERO BITS EXPECTED.

Figure 3-4. Input block diagram.

3.2.2 Implementation

Figure 3-5 shows *input* in more detail. The major elements are: data selector, coefficient memory, coefficient data path, address generator, multiplier-adder, overflow suppressor, test output selector, and initializer. All elements except the initializer are clocked by the **PHI** clock input. The initializer has a separate **iclk** clock input.

3.2.2.1 Data Selector

Each input, **d0**, **d1**, **d2**, **d3**, to the data selector passes through an acceptor, Section 3.1.4, and one flip-flop delay. The MUX selects the desired delayed input, which is delayed again. The select inputs for the MUX come from the DS register in the initializer. When the 11th and 12th bits of the input word are on the **qa** lines, these bits and the word count **wc** are loaded into RA, the read address register. The **test** input is used to simplify testing of the data paths, multiplier-adder, and overflow circuits. When **test** is TRUE, rather than CX and CY being loaded from the coefficient memory, delayed input **d0** is shifted into CX, **d1** into CY, and **d3** into CD. With **test** FALSE, the selected data input is shifted into CD, and ZERO into CX and CY as coefficients are shifted out.

3.2.2.2 Coefficient Memory

The coefficient memory is a 32-word, 22-bit CMOS SRAM using conventional 6-transistor memory cells. Static predecoders and row-decoders, P-channel bit-line clamps and prechargers, and a simple inverter sense-amplifier are used. The bit-line clamps and static decoders make this memory completely static. Similar circuits are used in the *delay* memory, Section 3.3. The memory is written from the initializer and obtains its address from the address generator.

3.2.2.3 Coefficient Data Path

The memory is read using the 11th and 12th bits of the selected data word and the word count as address. The offset and slope values obtained from memory are loaded into CX and CY just as the data word LSB arrives at the CD register output. These three values are then applied to the multiplier-adder along with the sync bit from CS. The logic at the input to CS separates sync and frame bits. The frame bit enters the F register and is shifted right after each sync bit. When a frame bit is present on the sync line, the word count is reset to ZERO. Taps on the CS register control loading of RA, reading the memory, and loading CX and CY.

3.2.2.4 Address Generator

The address generator contains a multiplexer to select either the Read or Write address, a word count generator, and gates to disable memory read and RA loading when **test** is ONE. When the

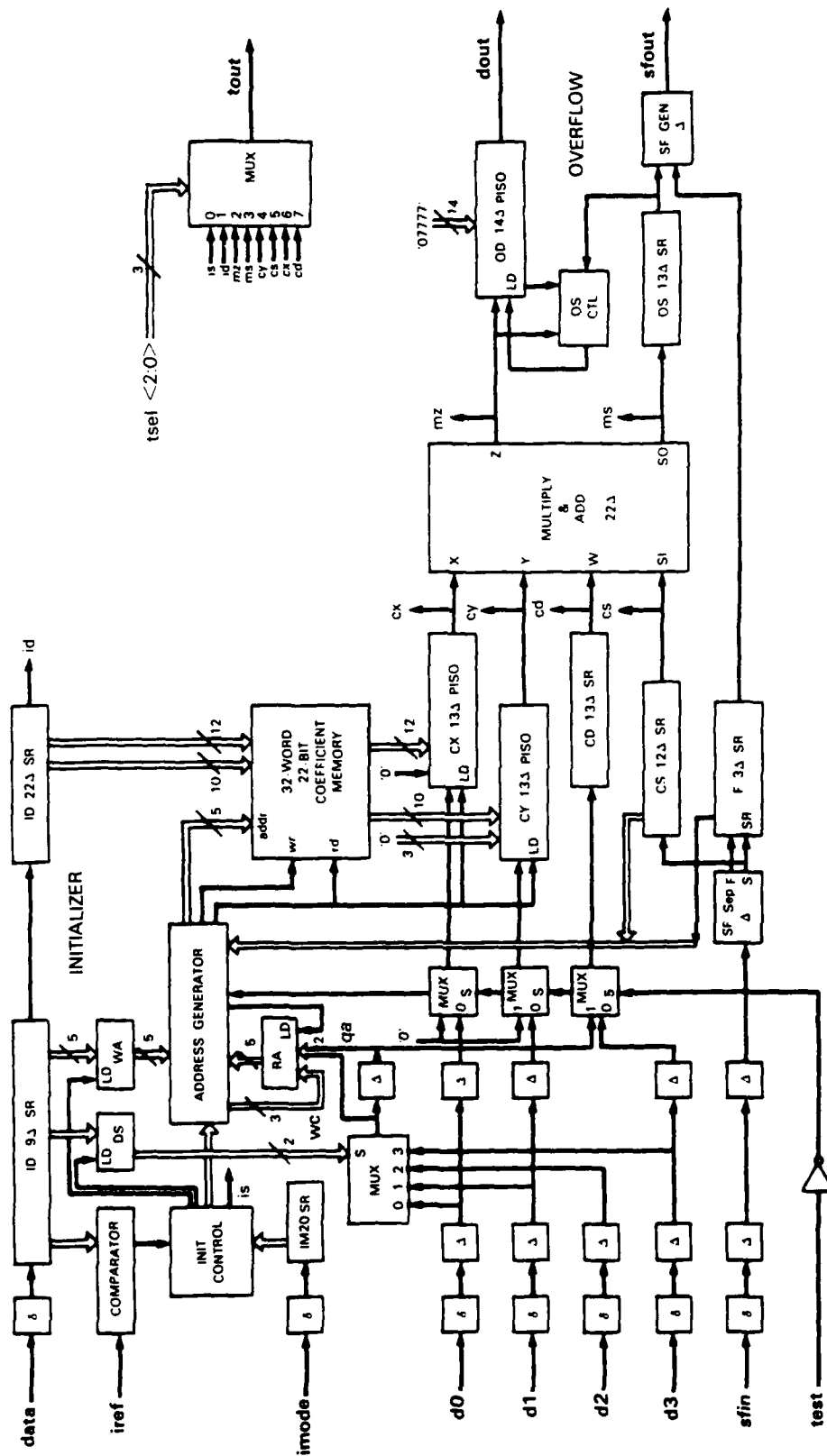


Figure 3-5. Input circuit implementation.

initializer is writing coefficients in memory, WA is used as the address. The word count generator output is either the current count plus 1 modulo 8, or 0 if the frame bit is ONE.

3.2.2.5 Initializer

The initializer consists of a 31-bit ID shift register, a 2-bit IM shift register, an address comparator, control, data select DS register, and a memory write address WA register. **Idata** and **imode** have input acceptors¹. Gates detect IM = 01 at the falling edge of **imode**, and IM = 10 at the rising edge. Seven bits of ID are compared either with a laser-programmed referent or, during wafer-probe testing, the **iref** input, which can make the referent either 000 or 177. The output of the address comparator is stored in the initializer select bit IS when IM = 01. Initializer data and a 1-bit opcode are also shifted into ID. If IM = 10 and the select bit IS is 0, nothing happens; however, if IS is 1, the opcode is examined, and the indicated operation performed. If op = 0, the data select register DS is loaded from ID. If op = 1, the write address register WA is loaded, and the addressed memory word is written from ID.

3.2.2.6 Multiplier-Adder

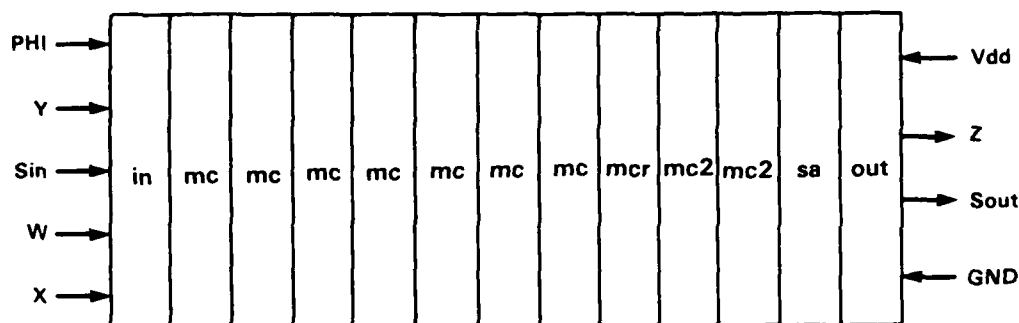
The pipelined serial multiplier-adder is a modification of the circuit used in an earlier WS circuit [8] which was derived from [9] and [10]. The differences from [8] are in number representation and the use of static flip-flops rather than dynamic storage in internal shift registers. As few other changes as possible were made, several circuits could have been a little simpler with more redesign. The result of one compromise is that the multiplier requires passage of one word for correct initialization. The circuit is used to multiply two 10-bit unsigned integers, Y and W , round the product to 12 bits, and add a 12-bit unsigned offset, X , to produce the output, Z . Specifically:

$$Z = \text{integer part of } \left(\frac{Y \cdot W}{256} + \frac{1}{2} \right) + X.$$

Or, W may be thought of as a number in the range $0 \leq W \leq (4 - 2^{-8})$ that multiplies Y to produce a rounded integer product.

The multiply-adder comprises 11 stages of pipelined serial arithmetic sandwiched between 2 terminating stages (Figure 3-6). The first 7 of 10 multiplier stages are the *mc* circuit of Figure 3-7 which stores one bit of coefficient, w_i , multiplies it by the multiplicand, y_i , and divides the product by 2. The 8th stage, *mcr*, is identical, except that the carry input to its adder is initialized to a ONE by a NAND gate, rather than a ZERO by a NOR, in order to add the 1/2 in the product and so effect a round. The final 2 multiplier stages use the *mc2* circuit, identical to *mc* except that it lacks the recirculation path for the partial-product, **PPo**. This eliminates the division by 2; over 2 stages it effects the multiplication of the eventual product by a factor of 4.

¹ Parameter bus inputs on other circuits do not have input acceptors. These would probably be removed since the parameter setting bus can operate at low speed.



127224-35

Figure 3-6. Circuits in the multiplier-adder.

Following the multiplier is a serial adder, *sa*. Its addend (X_i) has been delayed two clock cycles in each of the previous multiplier stages, and so is synchronized with the augend (PP_i). The input stage, *in*, comprises a phase-splitting clock driver and a few interconnect wires; the output stage, *out*, one driver each for the datum and synchronization signal.

3.2.2.7 Overflow Suppressor

The output of the multiplier-adder may exceed the 12-bit limit. Its output and sync are shifted into the overflow suppressor. If either the 13th or 14th bit of the data word are ONE, the OD register is loaded with 07777, the largest permissible value. The logic at the output of the OS register reconstructs the sync-frame signal from the outputs of OS and F.

3.2.2.8 Test Output Selector

To further simplify testing, the $t\text{sel} < 2 : 0 >$ inputs switch one of eight internal states to the **tout** output. This output must be used with care since it is not clocked.

3.2.2.9 Test Results

The *input* cell was fabricated by Mosis in Run M8BZ. There were 12 packaged devices received and tested with a partial test pattern file which tested all data paths using **tout**, **dout**, and **sfout**, but only initialized and read 2 words of the memory; 9 devices passed all tests at 1 MHz, 3 failed. Using only the functional outputs **dout** and **sfout**, 1 device was tested at a higher rate and operated correctly to 18 MHz.

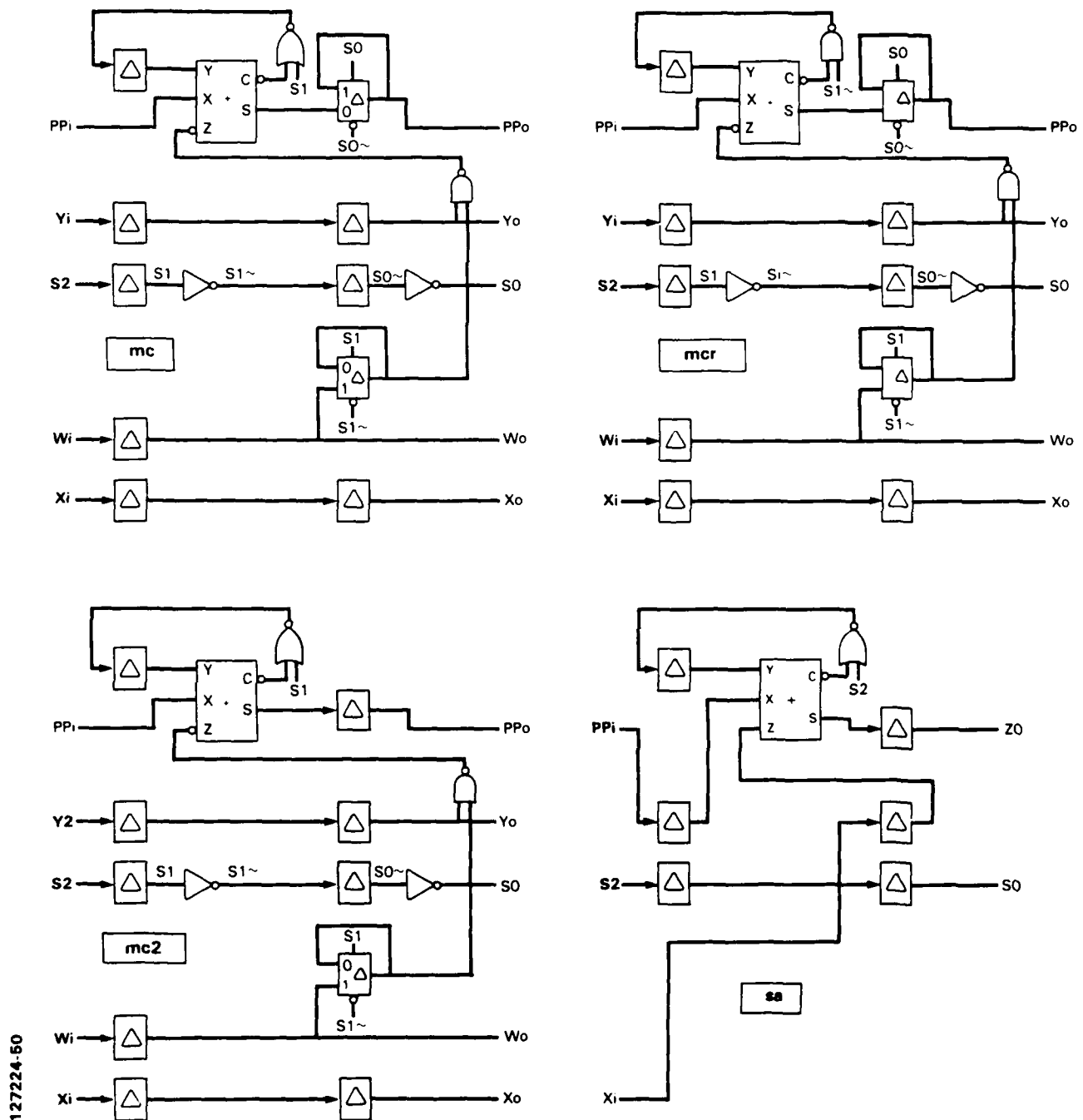


Figure 3-7. Multiplier and adder circuits.

3.3 DELAY

3.3.1 Function

The rate of rotation of the mechanically scanning detector array is such that a detector is looking at the same target space as a detector in the same row and adjacent column four sampling times previous. The signals from five detectors for the same target space are to be summed so that signals out of the *input* circuits must be delayed by different amounts before the summation. Delays in the five signal paths are $32n$ (where $n = 0, 1, 2, 3, 4$) word times. The delay circuit is realized as one 32-word delay and n circuits are cascaded. Logically, the delay is 14×32 clock periods, but a shift register implementation would have dissipated excessive power so a 32×12 static memory is used as a circular buffer. This circuit requires no parameters. Figure 3-8 is a block diagram of this cell.

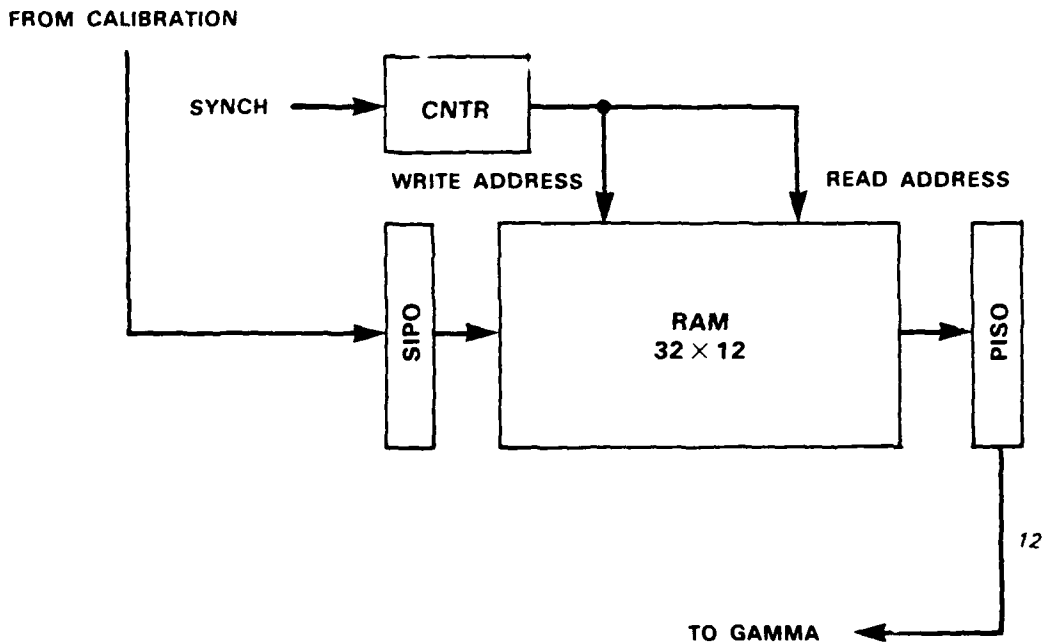


Figure 3-8. Delay block diagram.

3.3.2 Implementation

The *delay* unit is implemented with a 32-word 12-bit CMOS SRAM, a 14-bit serial/parallel-input data register D, a sync-frame separator and sync register S, an address register A, an address generator, and a parallel-input write register W, as shown in Figure 3-9. The SRAM is similar

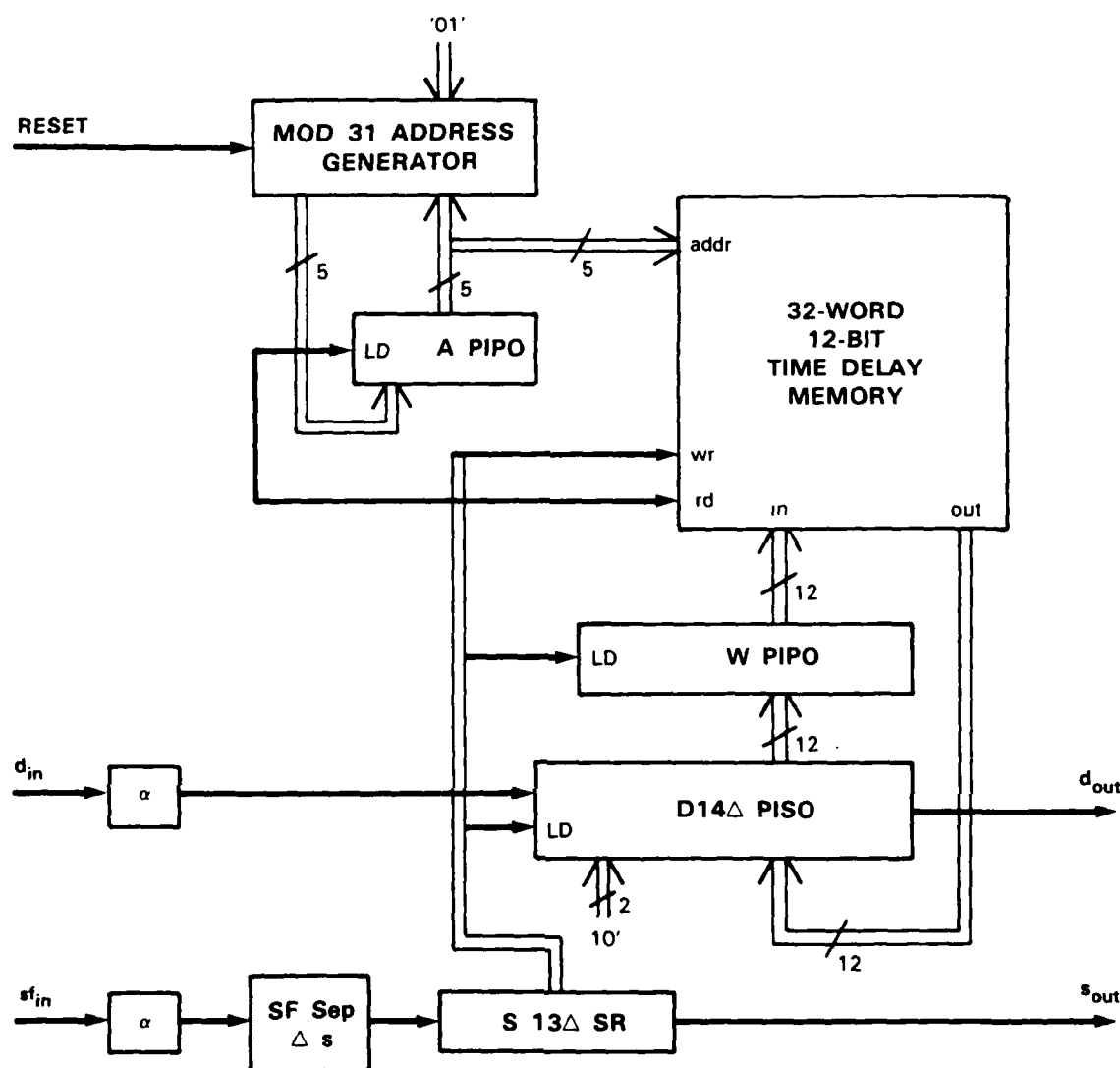


Figure 3-9. Delay circuit implementation.

to the *input* circuit memory described in Section 3.2. Input data are shifted into the D register, while delayed data previously loaded to D are being shifted out. As the MSB is shifted in, the new input word in D is transferred to W and replaced by the delayed word from the memory. During the next word period, the contents of W are written to memory, the address is incremented modulo-31, and the following delayed word read from memory. Because of the one-word delay in the D register, a modulo-31 count is required to achieve a 32-word delay. A reset input is provided to allow controlled startup of the address counter during testing.

3.3.3 Testing

The *delay* cell was fabricated by Mosis in Run M89O. Twelve packaged devices were received and tested with a test pattern file which tested the D and S registers, wrote and read alternating 10s and alternating 01s in all words, and checked 31 random words for proper delay. All devices passed at 16 MHz. One device, tested at higher rates, operated correctly at 20 MHz. At 16 MHz, using the above test pattern, the average I_{dd} of one circuit was 4.8 mA, which dropped to 0.28 at 1 MHz.

3.4 THRESHOLD

3.4.1 Function

The *threshold* and *average* cells perform the gamma circumvention and TDI summation. Briefly, in each set of five signals which are to be averaged together, any signal which is larger, by some constant times the standard deviation, than the average of all five is eliminated from the averaging process. The reader is referred to [2] for further details. In this cell the sum of the five signals is generated and divided by 5. The square root of the average, which for a Poisson process is the standard deviation, is computed, multiplied by a parameter, and the result is added to the average to create a threshold. Mathematically:

$$thresh = \hat{\lambda} + k\sqrt{\hat{\lambda}}.$$

The only parameter for the *threshold* cell is the 5-bit quantity k . Figure 3-10 is a block diagram of this cell.

3.4.2 Implementation

Figure 3-11 gives a detailed view of the implementation of the *threshold* circuit. It can be broken into two parts: The first, much the larger, performs the computation of the preceding equation. The second loads the single 5-bit parameter G ; except for the detail of the loaded parameter it is identical to that of the *filter* circuit, described in Section 3-6. The computational circuit has five data inputs, **I0** through **I4**, an accompanying timing input, **Sin**, and a single clock, **PHI**. There is one data output, the threshold value **Thresh**. The estimated mean is to be computed as

$$\hat{\lambda} = \frac{1}{5} \sum_{n=0}^4 I_j$$

and k as

$$k = \frac{G}{32}.$$

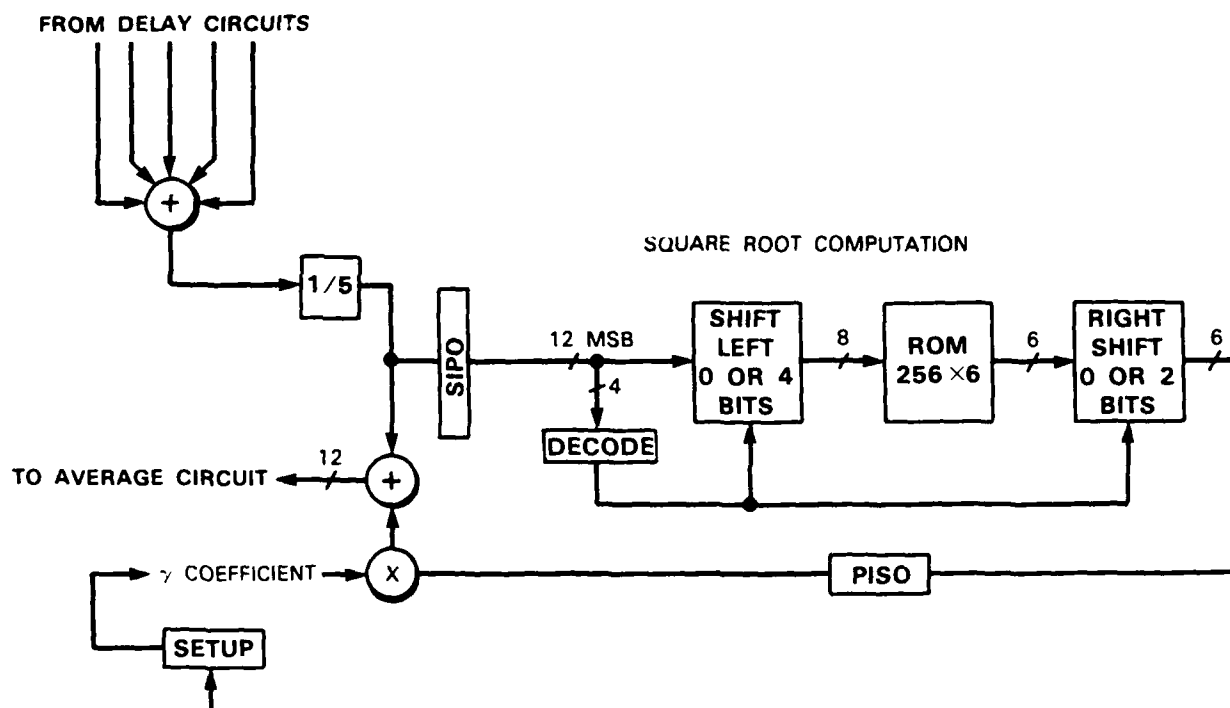


Figure 3-10. Threshold block diagram.

Arithmetic is performed on positive integers, $I_j < 2^{12}$. Division, always by a power of 2, is effected by right shifting; at appropriate points a rounding term is added. In the diagram, subcell *add5* sums the inputs (to S), and with multiplier *mk6*, in effect, multiplies by $1/5$ to give the mean P . Subcell *sqr* takes the square root R in a 256-word by 6-bit ROM and *mk5* multiplies it by G to give the product T , which is then added to the mean to give the output limit **Thresh**:

$$S = \frac{1}{4} \left(\sum_{n=0}^4 I_j + 2 \right)$$

$$P = (51 \times S + 32) / 64$$

$$R = \sqrt{16P}/4, \quad P < 2^8$$

$$= \sqrt{16(P/16)}, \quad P \geq 2^8$$

$$T = (R \times G + 16) / 32$$

$$\text{Thresh} = P + T.$$

Output **Thresh** can be as large as 13 bits: $\text{Thresh} \leq 2^{12} + 2^6 - 20 = 4140$. There are four operational output test points: **TP1**, **TP2**, and **TP3** represent states S , P , and T , with delays of 7, 19, and 35 cycles from the input, respectively; **TP4** is the synchronization output, simultaneous with **Thresh**, 46 cycles after the input.

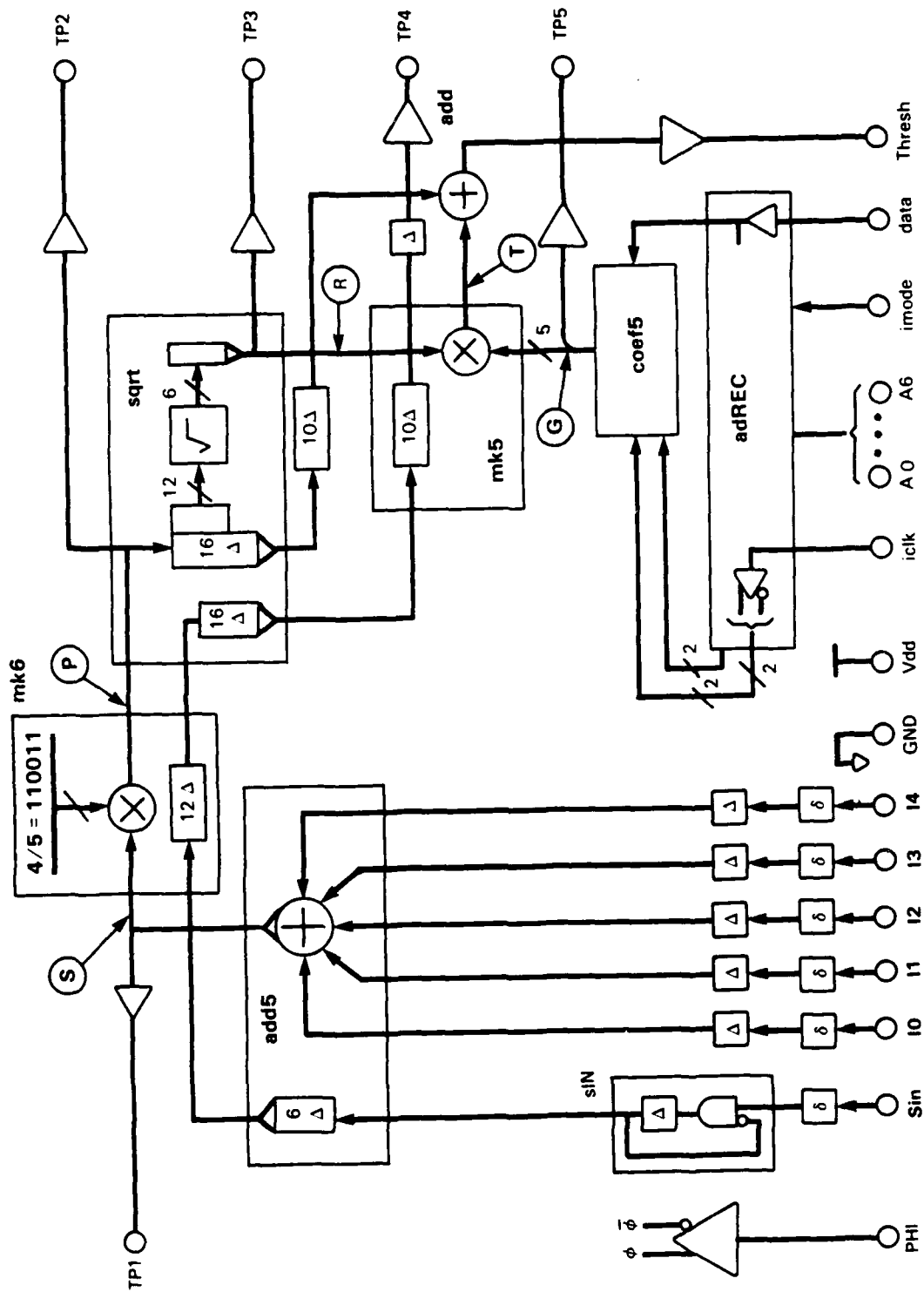


Figure 3-11. Threshold circuit implementation.

3.4.2.1 Five-Input Adder

The adder, *add5* (shown in Figure 3-12), is designed to produce the most precise possible 13-bit sum of five 12-bit positive integers, i.e., the one with the least error in roundoff. It comprises five identical serial 1-bit adders, associated delays, and a switch to send the final carry bit of the fourth adder to the fifth for summation with the 14-bit sum and a rounding bit. The 1-bit adder stage is similar to that used in the matched-filter circuits. The first three adder stages produce a 14-bit sum of inputs I_0 through I_3 in a straightforward manner. Operation of the last two stages is a bit more complicated. It may be better understood in a representation as a two-stage parallel or array adder, as in Figure 3-13. Here the a_n are the 14-bit intermediate sum and b_n the fifth (I_4) input. First-stage sum bits, except for the least significant, are added to a single rounding bit in the second stage, along with the most-significant carry. A 13-bit output is extracted from the final adder by discarding the second-stage LSB. (In a serial adder, of course, the selection of output bits is effected not by wiring, but by time shifts.) Notice that the second-stage carry input C_i appears undefined. Actually, it is the carry output C_o of a similar circuit. (In a serial adder, this is the residual carry from a preceding sum.) For the possible range of inputs it will always be zero: $(2^{14} - 4) + (2^{12} - 1) + 2 < 2^{15}$.

3.4.2.2 Multipliers

The 5- and 6-bit serial multipliers are similar to those used in the 10-bit multiplier of the input stage, with two changes. First, because they use fixed coefficients, shift registers for serial-to-parallel conversion and storage are unneeded. Second, their arithmetic has been customized for positive integers, making it possible to use a single type of stage in all locations.

3.4.2.3 Square-Root Circuit

The most direct method to generate the required square root would be a ROM with a 12-bit address for 6-bit words. Instead, a ROM with only an 8-bit address was used. Inputs $P \geq 2^8$ were divided by 4 before being used as a memory address; those $P < 2^8$ were applied directly, and the resulting ROM output divided by 4. The result is a good approximation of the true square root for both large and small inputs. The ROM is derived from the SRAM used elsewhere. The SRAM memory cell and access transistor pair are replaced by a common-source transistor to either the ZERO or the ONE bit-line depending on the stored bit. The ROM uses 32 rows and 8 bit-line pairs for each digit. A completely static radiation-hardened ROM is created with cross-coupled digit clamps and static predecoders, row-decoders and column-decoders.

3.4.2.4 Output Adder

The final adder is similar to the *ar* adder of the matched filter, but since no rounding is required and inputs are positive integers, the NAND and NOR gates are eliminated.

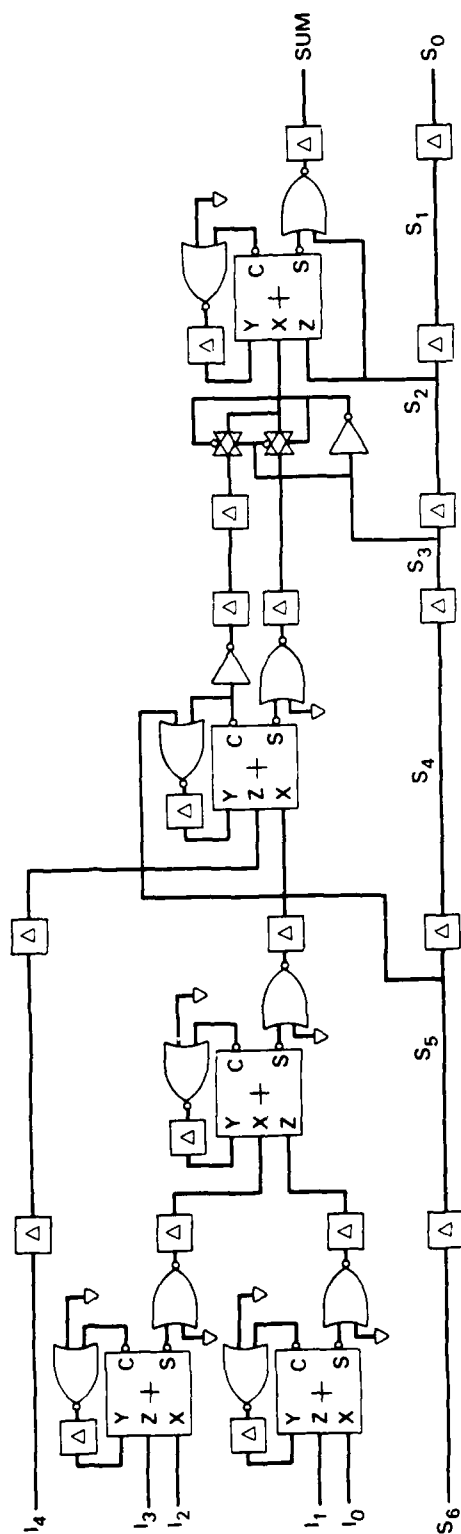


Figure 3-12. Five-input adder circuit.

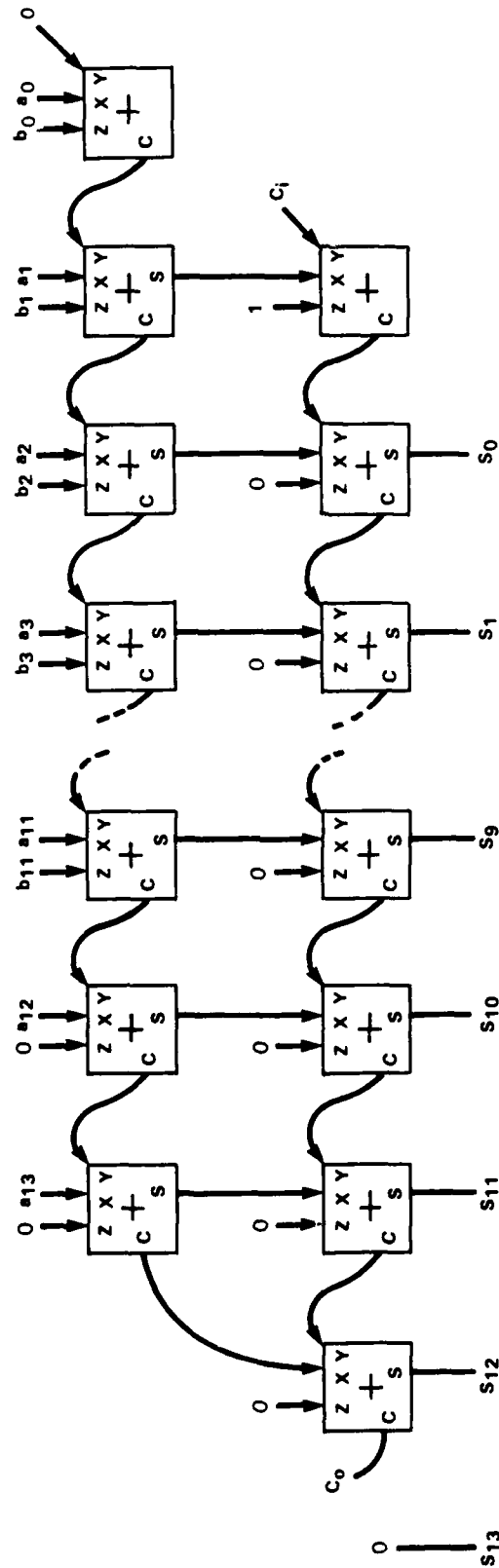


Figure 3-13. Parallel description of two adder stages.

3.4.2.5 Synchronization Circuit

The *threshold* synchronization circuit consists only of a delay and gate to strip off the second half of a double-width marker pulse.

3.4.3 Testing

Testing philosophy is to do an exhaustive test for nodes stuck in 1 state and operate with random inputs to reveal high-frequency limitations. The parameter-loading circuit is tested in 10 segments of 13 loading cycles each. An arithmetic test exercises all cell arithmetic circuits except *sqr*t in 18 14-bit words associated with 3 sets of parameters: 4 words for the 5-input adder, 2 words for the 6-bit fixed-coefficient multiplier and the output adder, and 12 words for the 5-bit multiplier, including 2 sets of 4 null words each to complete computations with 1 value of **G** before the next is supplied. *Sqr*t is exhaustively exercised with 257 words. With a random test of 275 sets of integers, a complete test requires 8116 clock cycles.

Twelve packaged devices were received from a fabrication in MOSIS run M89O. All were functional at 10 MHz, and the three tested beyond that rate operated at 20 MHz even with the supply dropped from 5 to 4.5 V. At $V_{dd} = 5$ V, the mean output transition times into 50 pF loads were 32 ns. Mean delay for signal outputs relative to clock **PHI** transitions was 42 ns for rising transitions, 50 ns for falling; for parameter-loading test-point output **TP5** relative to **iclk**, 49 ns for rising transitions, 54 ns for falling.

3.5 AVERAGE

3.5.1 Function

Figure 3-14 is a block diagram of the *average* circuit. Its inputs are the same five signals as are input to the *threshold* circuit and the threshold value calculated in *threshold*. The five signals are delayed by the latency of *threshold*. If a signal is larger than the threshold value, then a ZERO is forced onto its adder input line. The sum of the five signals is divided by N, the number of nonzero adder inputs. There are no parameters in this circuit.

3.5.2 Implementation

The *average* cell of Figure 3-15 is a signal-processing circuit intended to compute

$$DATA_o = \frac{1}{n} \sum_{I_j \leq Thresh} I_j \text{ where } n = \text{number of } I_j \leq Thresh.$$

As implemented, it can be described by

$$L = \frac{1}{4} \left(\sum_{I_j \leq \text{Thresh}} I_j + 2 \right)$$

$$\text{DATAo} = \frac{L \times M + 512}{1024}$$

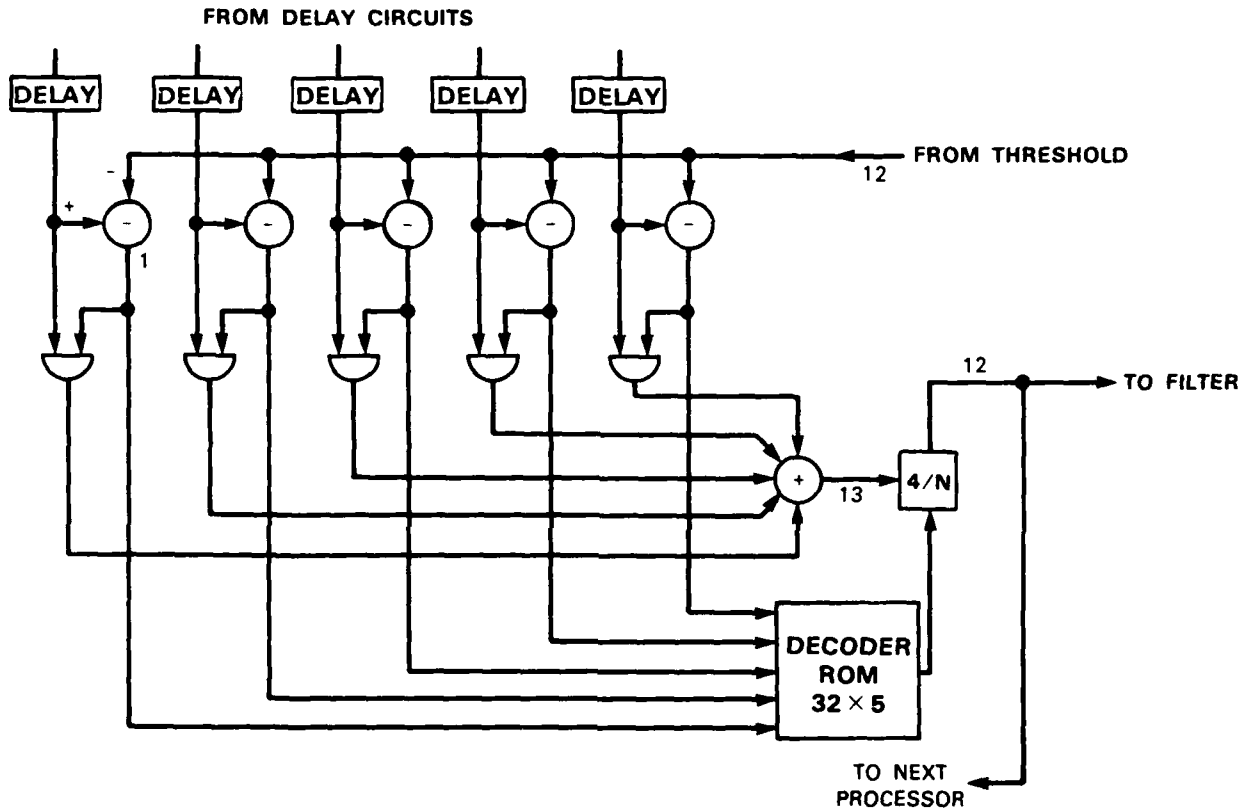
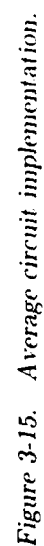


Figure 3-14. Average block diagram.

where rounding terms are included and the averaging division is approximated by multiplication by a coefficient **M** as shown in Table 3-1.

The sum **L** is produced by a high-precision 5-input adder of the type used in the *threshold* circuit, with any inputs exceeding **Thresh** set to zero by the input subtractors. The final product **DATAo** is produced by a multiplier identical to that of the *input* cell, except that it uses a 12-bit coefficient, rather than 10.

The one unique circuit of this cell is that used to produce coefficient **M**, Figure 3-16. Its *tally* circuit, a fully complementary implementation of a textbook circuit [11], makes one of 5 lines high, according as the number of zeros in the 5 subtractor sign bits. These lines operate a 5-input selector for the **M** signals of the table. Note that in the table, 2 of the signals have single ONE



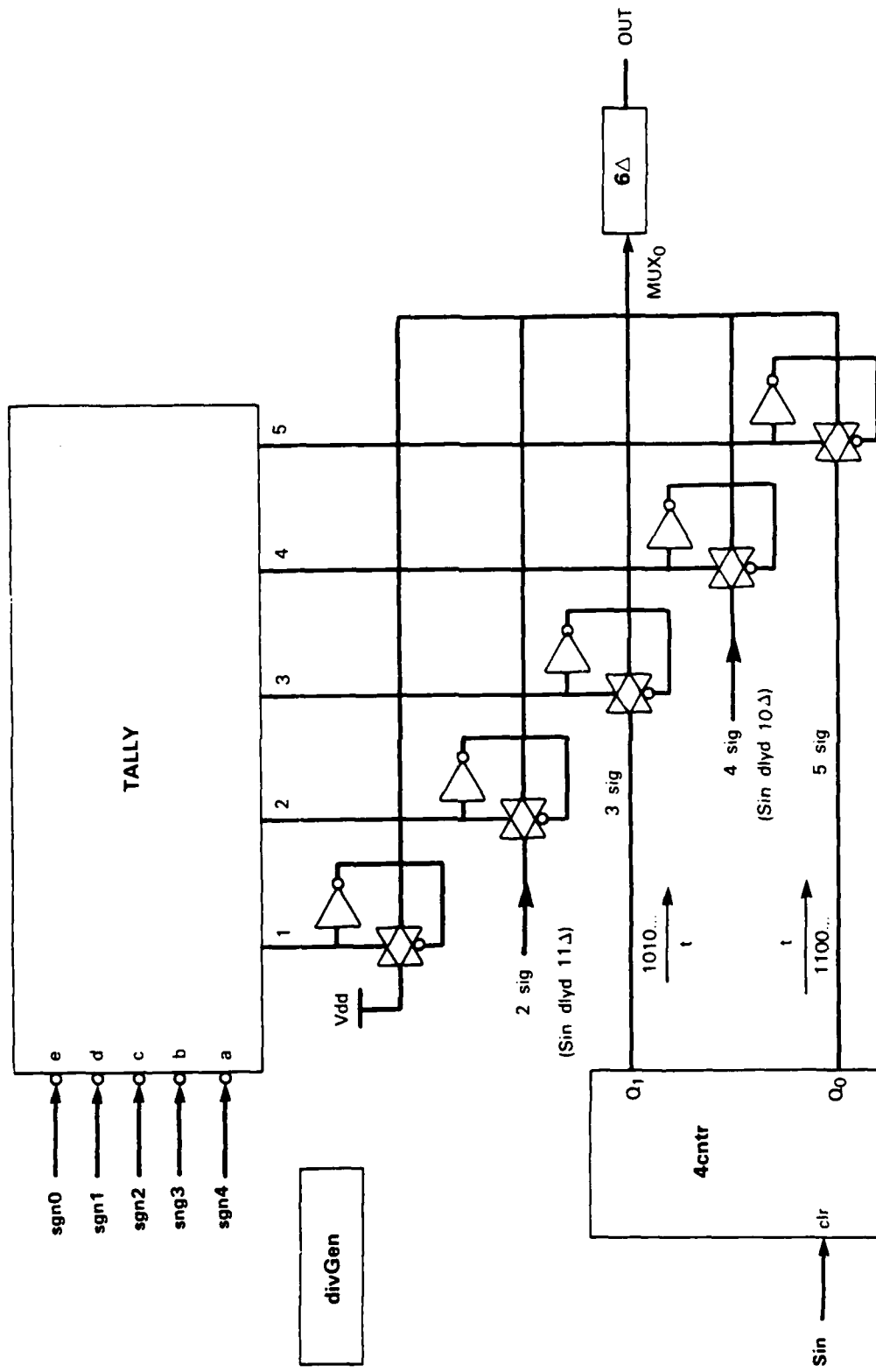


Figure 3-16. Average coefficient generator.

bits; these are tapped off the synchronization channel. Because the multiplier responds to only the 12 least-significant of the 14 coefficient bits, the 2 most significant are "don't care." or **X** in the table, which means that the signal for $n = 1$ can be DC, and the remaining two can be periodic, generated by a resettable scale-of-four counter.

A final subtlety about the coefficient-generation circuit was not recognized when the preliminary circuit design was done: Switching between 2 different coefficient words can occur at any of 3-bit positions, preceding (in time) either of the 2 don't care bits of the first word, or following the entire word. Because of the sluggish combinatorial logic of the tally and selection circuits, at high clock rates switching that is done following the word may not be complete at the time of the LSB of the following word. It is better to adjust control timing so switching is done between the don't care bits.

The data path for the synchronization signal, **Sin** to **Sout**, is similar to that of cell *filter*. Input synchronization is simultaneous with the data inputs I_j ; output, with **DAT0o**. A double-width marker is stripped from the input, and reinserted at the appropriate time in the output: the internal synchronization pulse is a single bit.

There are two output test points: **TP1** samples the multiplicand input of the output multiplier. **TP2** the coefficient.

Throughput delay from the data to **Sout** and **DAT0o** is 92 cycles; test-point outputs are 24 cycles earlier. Input **Thresh** must be delayed 46 cycles relative to the data.

3.5.3 Testing

The test philosophy was to make an exhaustive search for nodes stuck in one state, and then add operation with random inputs to reveal high-frequency limitations. The performance of the input subtractors must be inferred from generated coefficients observed at **TP2**. An exhaustive test of both difference and borrow outputs of the combinatorial subtractors for all 8 input combinations and for borrow initialization required 17 sets of input words, including many falling outside the normal 12-bit data values. A test of the tally circuit of the coefficient generator required 31 input data sets. These tests were combined, so that the total number of word sets was also 31. The test for the 5-input adder was that used in the *threshold* cell. The 12-bit multiplier was tested by 8 input word sets: 3 for coefficient storage, 1 for carry initialization, and 4 for an exhaustive test of the combinatorial adders. To these systematic tests were added 250 sets of random inputs, chosen so that at least 1 datum of each set would not be less than the input **Thresh**, for a complete test in 4091 clock cycles.

There were 12 packaged devices received from a fabrication in MOSIS run M88F. Of these, 10 functioned at 12 MHz, but all failed at outputs **TP2** and **DAT0o** at 13 or 14 MHz. Failure was a consequence of the less-than-optimum accommodation of delay in the logic that produces the final "divisor." Since this can be corrected by a simple wiring change, a better measure of circuit capability is obtained from output **TP1**. There was functionality to 20 MHz even with the supply reduced to 4.5 V.

3.6 FILTER

3.6.1 Function

The matched filter and detector is a digital filter to perform a separable 4×4 two-dimensional convolution on data from an array of photodetectors and compare the result with a detection threshold. It is shown in simplified form in Figure 3-17. The matched filter has two input data streams, i from the average circuit of this processor, and x from another processor but initially it is simpler to consider only stream i , and assume the multiplexer switches make permanent connection to it. Data for one processor comes from a raster scan of an eight-row by five-column array of photodetectors, but the data from the five detectors in each row have been averaged by the *average* cell. The detectors are oversampled by a factor of four. Data from filter-input unit-delay taps are multiplied by coefficients and added for a vertical convolution, and the results in turn delayed, multiplied by coefficients, and added for a horizontal convolution. The result of the final convolution is added to a threshold; the sign of the output indicates detection status. The second input data stream, x , is necessary at the edge of the raster scan. In convolving data from a detector column, the filter deals with the eight detectors associated with its particular processor—the i or “intrinsic” data—and the top three adjacent detectors, associated with another processor—the x or “extrinsic” data. In a sequence of eight 4×4 convolutions, the first five involve only i data, and the next three involve progressively one, two, and three rows of x data. The multiplexers of Figure 3-17, under control of a scale-of-eight sample counter, provide the necessary switching between data streams.

Mathematically the filter circuit is described by

$$z[n] = t + \frac{1}{256} \sum_{j=0}^3 H_j \sum_{k=0}^3 V_k \{ s[5+k-(n \bmod 8)] i[n-k-8j] \\ + s[(n \bmod 8) - 4 - k] x[n-kj] \}$$

where the unit step function,

$$s[m] = 0, \quad m \leq 0, \\ = 1, \quad \text{otherwise,}$$

effects the periodic switch between data streams. In the equation $i[m]$ and $x[m]$ are the m th input data; H_j is the j th horizontal coefficient of the convolution kernel and V_k the k th vertical one; t is the detection threshold; and $z[n]$ is the n th filter output. The input data are 12-bit positive integers, the coefficients 2 bits, and the threshold and output 14-bit signed integers. Products and sums within the summation are rounded to 12 bits.

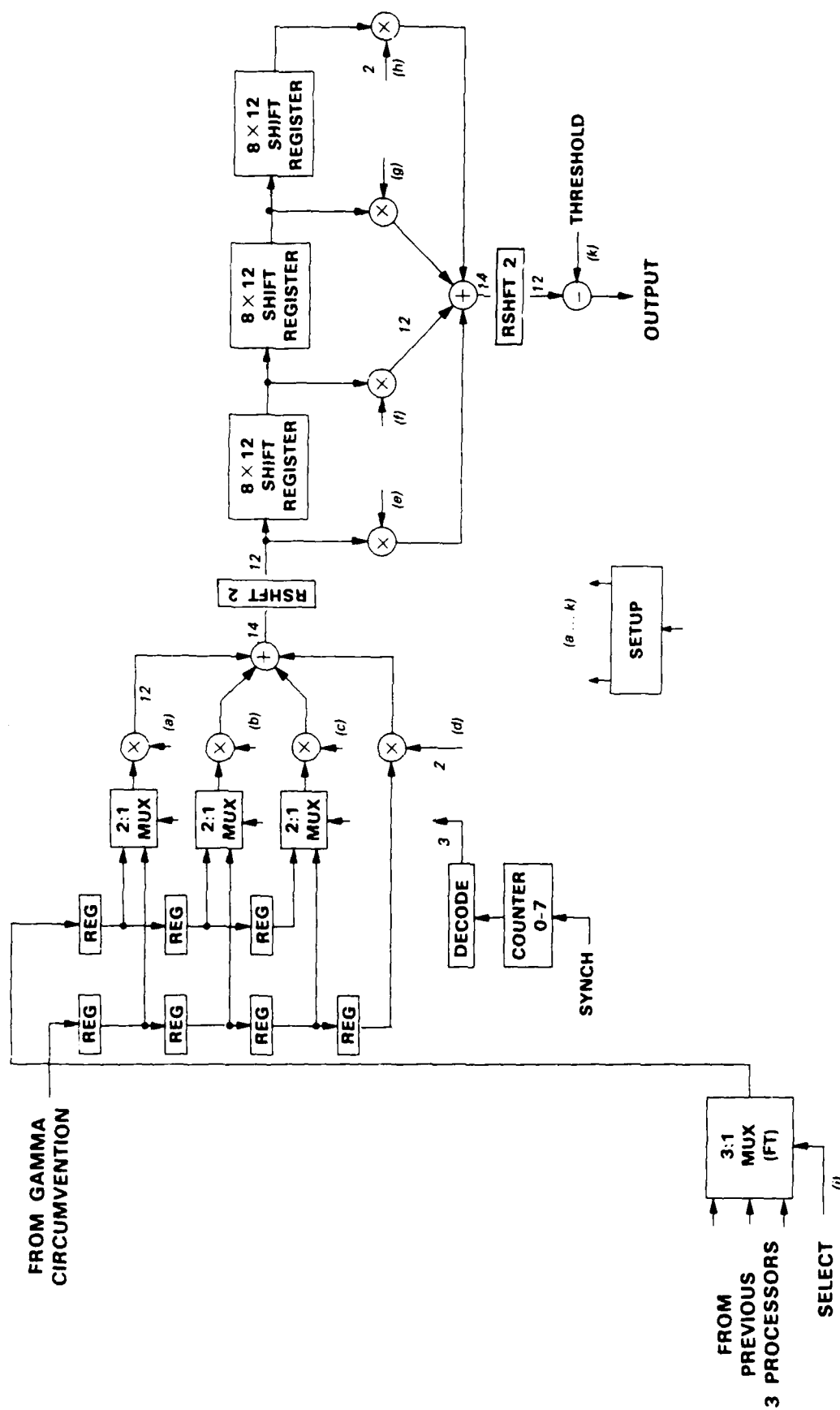


Figure 3-17. Filter block diagram.

3.6.2 Implementation

Figure 3-18 shows the *filter* circuit in more detail, in particular, the delays through the circuit. The descriptive equation shows a filter output dependent on the previous 27 input words, for a span of $14 \times 27 = 378$ clock cycles. To this is added a 17-cycle delay in filter circuits for a total latency of 395 cycles. (For the *filter* circuit in isolation, the acceptor circuit might also be considered to have a delay of one-half cycle.)

The filter cell can be broken into two parts. The first is the circuit to perform the mathematics. The second is that used to load and store the fixed parameters. In Figure 3-18, the latter is found at the bottom center, in subcells *setup*, *adREC*, and a portion of *cmpThr*; the former comprises everything else. The matched filter performs serial arithmetic under control of the single-phase clock **PHI** and word-timing or synchronization input **Sin**. *Intrinsic* data are taken from terminal **i**: *extrinsic* from **X0**, **X1**, or **X2**, dependent on the value of the preloaded address **muxAD** to the three-input multiplexer. Data output is at terminal **OUT** with a corresponding synchronization signal **Sout**. Input data are positive integers in 14-bit words, presented with the LSB, 0, first. The two MSBs must be ZERO to serve as guard bits. Word boundaries are denoted by a ONE in the bit 0 position of the synchronization signals. Every eighth word is marked with a double-width synchronization pulse. Output data are 2's complement integers. The convolution coefficients, **H_j** and **V_k**, and threshold, **t**, are preloaded as parameters. The subcircuits will be discussed in some detail.

3.6.2.1 Quad Multiplier and Adder

The quad multiplier and adder subcell (*qm0* and *qm1*) of the matched filter consists of four rounding multipliers for serial positive-integer multiplicands Y_n and 2 bit parallel coefficients W_n and a tree of three rounding adders to sum the products. Mathematically:

$$Z = \frac{1}{16} \sum_{n=0}^3 Y_n W_n.$$

A block diagram of the quad multiplier cell, *qm2a* is shown in Figure 3-19. Explicit in this diagram are four multiplier stages, three serial adders, and a synchronization delay chain. Each multiplier performs:

$$P_i = \frac{Y_i W_i + 2}{4}$$

and each rounding adder:

$$S_i = \frac{X_i + Y_i + 1}{2}$$

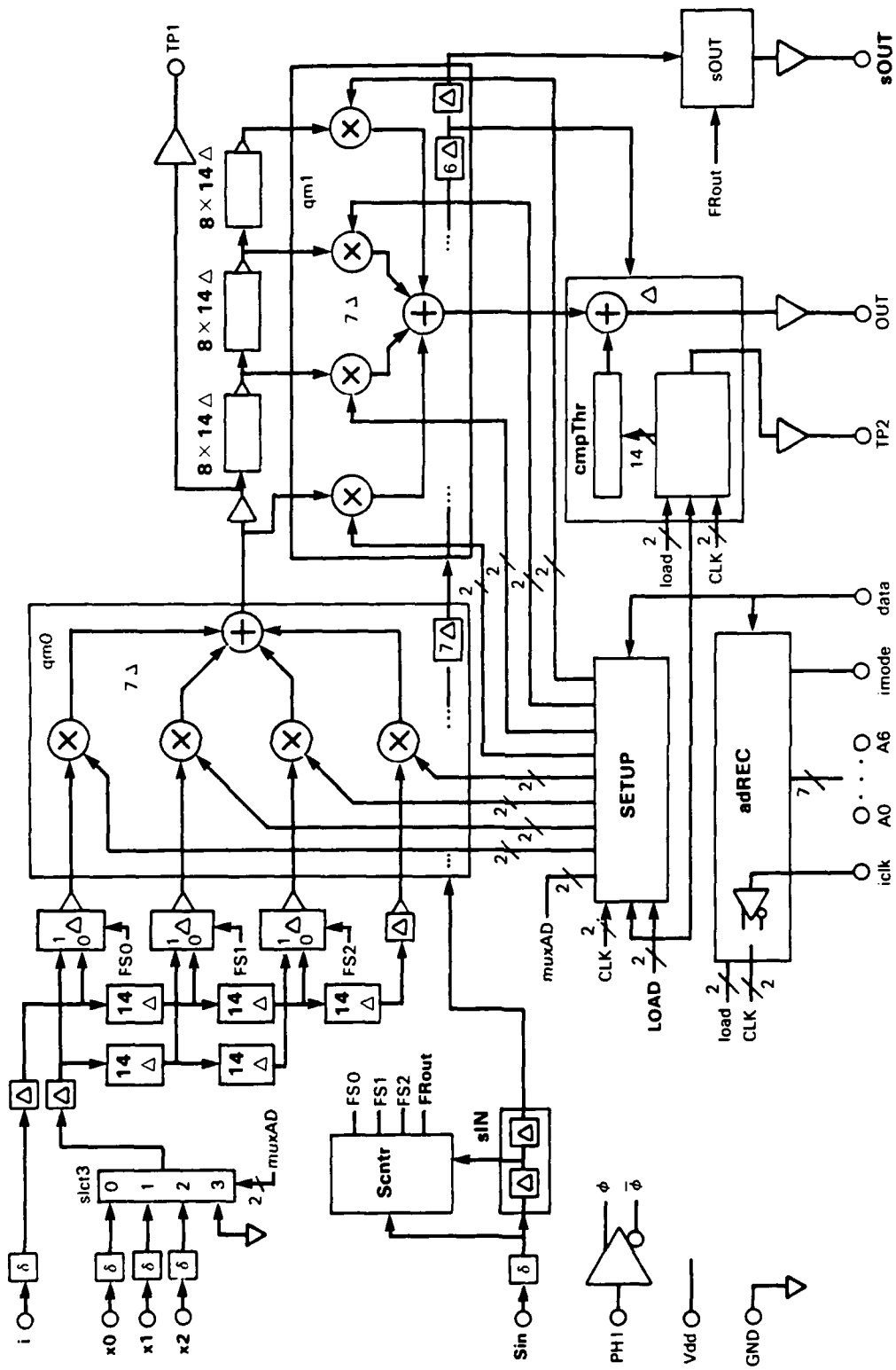


Figure 3-18. Filter circuit implementation.

Figure 3-19. Quad multiplier-adder circuit.

Because of the rounding associated with each addition, the output sum depends on the order of product terms. For example, assume products $Y_n W_n = 0, 2, 4, 6$. If applied to $qm2a$ in that order, the sum is $Z = 2$, but if the order is $0, 6, 4, 2$ the sum is $Z = 1$. This order dependence is unlikely to be significant in filter operation, but it does require special care in simulation programs used to generate test vectors.

Figure 3-20 is a circuit diagram of the multiply-and-round stage, $m2r$. Partial products of the serial multiplicand Y_i and the parallel coefficient are applied to the inputs of the combinatorial adder (Figure 3-21), with the more-significant coefficient bit delayed one clock cycle. Adder carry output is recirculated to the carry input through a NAND gate that, in effect, supplies the 2 needed for the rounding operation. Adder sum output becomes stage product output \mathbf{Pr} after the NOR gate truncates the two LSBs and a one-cycle delay effects a new bit alignment with the synchronization delay chain.

A circuit diagram of the add-and-round stage, ar , is shown in Figure 3-22. Again, there is a combinatorial adder with recirculating carry that can be forced to a ONE, but this time with two

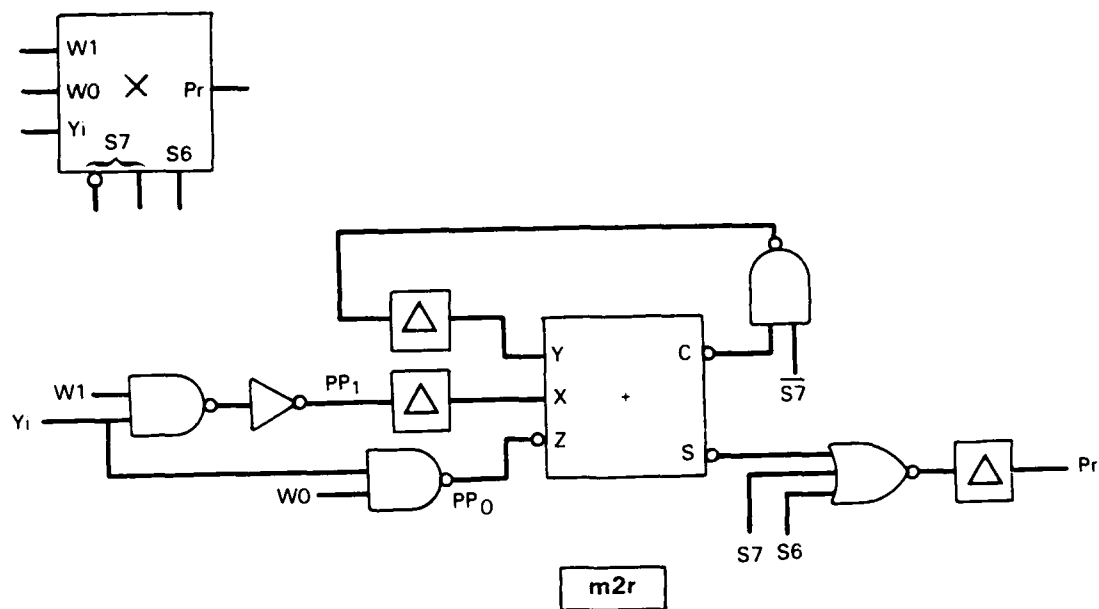


Figure 3-20. Multiply-and-round stage.

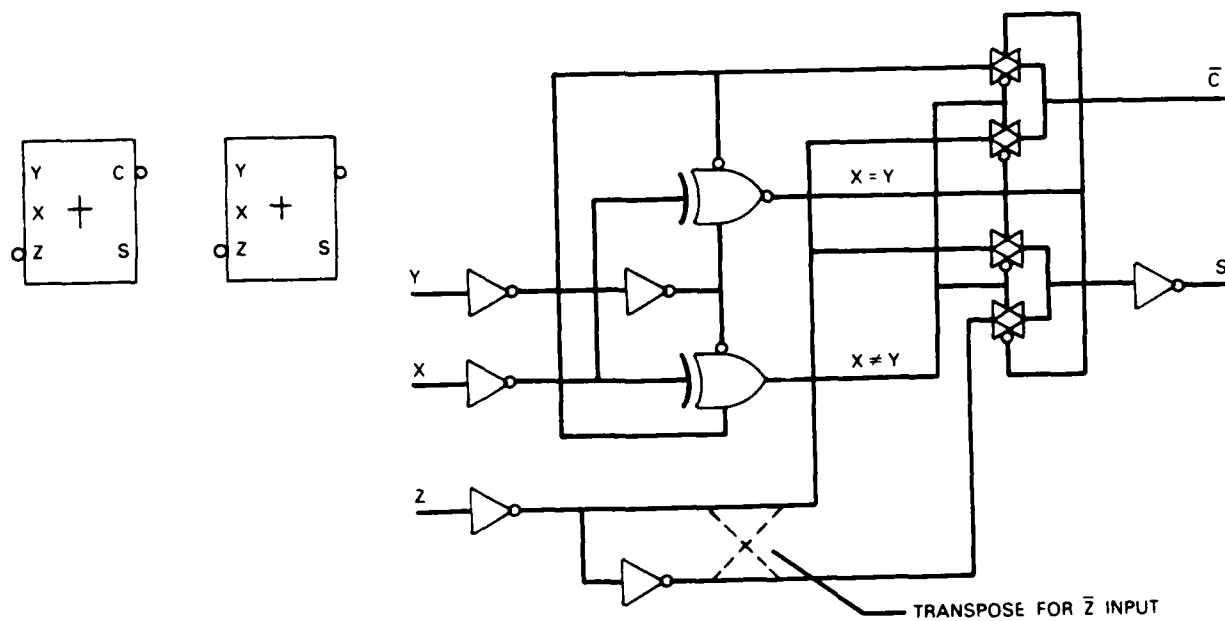


Figure 3-21. Combinatorial full adder.

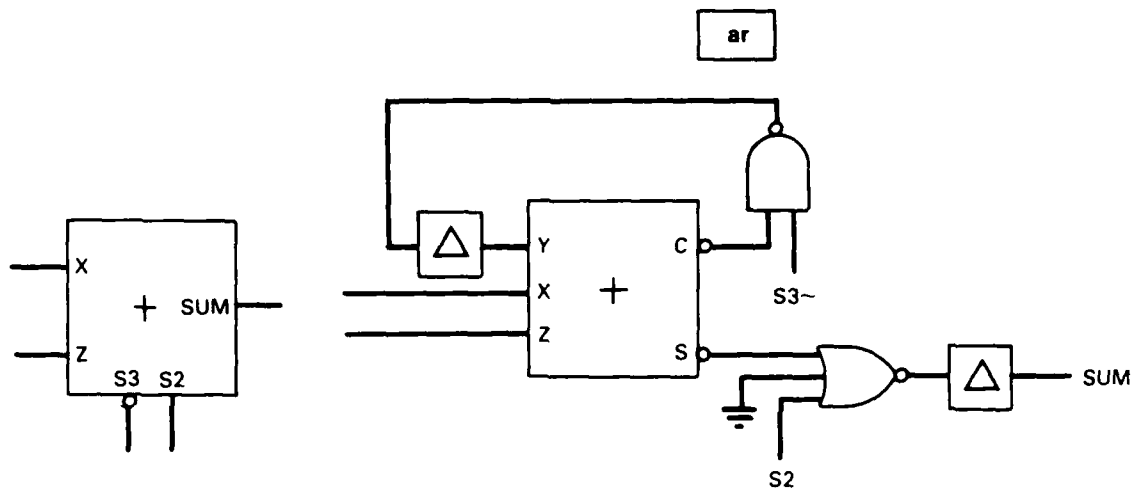


Figure 3-22. Add-and-round stage.

simultaneous addend and augend inputs and a two input NOR to set a single output bit ZERO.

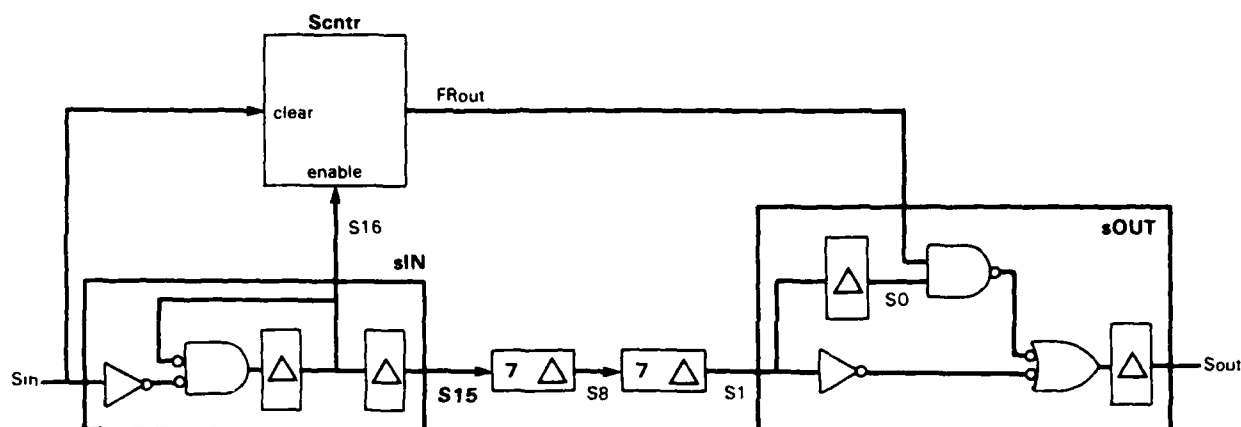
The maximum value of the quad-multiplier-adder output is 3071 so the two MSBs are always ZERO.

3.6.2.2 Threshold Comparator

The threshold value is loaded into *filter* from the global parameter bus. Once each word time the threshold value is parallel-loaded into a shift register and then serially transferred to an adder like the one in Figure 3-21, in synchronization with the output data from the filter.

3.6.2.3 Synchronization

The synchronization circuit of the matched filter is shown in Figure 3-23. An input subcell *sIN* and word counter *Scntr* are both driven by the *filter* synchronization input. *sIN* strips off the double-width frame pulse and applies the single-width pulse to the counter as an **enable** to advance the count if **clear** is ZERO, or reset the count if **clear** is ONE. Therefore, the double-width pulse causes a reset. The internal synchronization pulse is delayed one more cycle (*S15*) and passed through the two quad multiplier-adders. Then in subcell *sOUT*, it is given a final delay to bring it into step with filter output and, if a frame pulse is indicated, widened to two cycles. Counter outputs **FSn** applied to the two-input multiplexers progressively switch extrinsic signals to the inputs of the vertical quad multiplier.



127224-46

Figure 3-23. Synchronization circuit.

3.6.2.4 Parameter Circuits

With **imode** TRUE (3-18), serial **data** are clocked by **iclk** into the an 8-stage shift register: with **imode** FALSE, they are held in place by local recirculation. An address referent is set either in restructurable links by connection to **Vdd** or **GND**, or by levels applied to probe pads **A0** through **A6**. The stored serial address is compared to the referent by exclusive-or circuits and a 7-input NAND to generate the **load** level for the parameter storage register. The 32-parameter bits for the matched filter are clocked into a serially loaded register under control of **load**. Eighteen are stored in subcell *setup*, the remaining 14 in *cmpThr*. For test purposes the final bit is brought out to a pad, **TP2**.

3.6.3 Testing

3.6.3.1 Test Design

The emphasis in the design of functional tests for the matched filter was to make an exhaustive test for circuit nodes stuck in one state. The key words, exhaustive and stuck, require qualification. As far as possible, every circuit node is forced into both states in such a manner as to ensure that the consequences of these states are visible at a cell output. For some nodes this is impossible. For one thing, cell logic is replete with transfer gates, parallel N- and P-channel transistors with complementarily driven gates. If the drive of one transistor happens to be stuck at a potential that prevents conduction, the circuit will still function, although it may be a bit slow. One is, therefore, forced to add that the test is on *accessible* nodes.

The stuck-at tests tend to involve a few significant inputs in a sea of memory-clearing nulls, and, particularly at high frequency, may not be the most stressful. Therefore provision has been made for tests that involve random data inputs with, it is hoped, strategically chosen values of the filter parameters, coefficients, and comparison threshold.

Some general rules were adopted to keep generation of vectors for test of the cell simple and systematic. The parameter-loading circuit is tested only in segments of 40 cycles of its clock (8 for address, 32 for data). The filter itself is tested only with standard data words: 14 bits, with only the 12 least significant nonzero. Furthermore, the words are supplied only in frames of 8. The 2 circuit clocks operate in mutually exclusive fashion, one inhibited when the other is active. These rules limit test flexibility, and sometimes lengthen test duration, but it is believed they do not impair test rigor.

The remainder of this section on test design is concerned with the complete stuck-at test. The overall strategy follows. The parameter-loading circuit is tested first. Then signals are passed along cell data paths through the various switching and delay circuits, with minimal arithmetic operations. Finally, the arithmetic circuits are individually exercised in a combinatorially exhaustive manner.

Parameter-Loading Circuits Two parameter sets with matching parallel address referent and serial address are applied, the first with a bit pattern 1010..., the second with the complementary 0101... pattern. Then seven consecutive parameter sets are applied, each with the serial address mismatched by a different single bit. Finally, one more parameter set is applied with serial address matched to the referent. Data from the first parameter set are read out at **TP2** as the second is applied; from the second, as the last is applied.

Data Paths Data paths are tested by propagating an isolated ONE along some filter circuit paths with filter parameters set to evoke specific outputs. Four sets of vectors were devised for this purpose.

Arithmetic Circuits The test strategy for each of the arithmetic subcells of the matched filter is to ensure that its combinatorial full adder receives all eight possible input combinations; that all set and clear operations of special bits are performed; and that the results are visible at some filter output despite rounding operations in any subsequent circuits.

3.6.3.2 Test Implementation

Vector files for testing the packaged circuit were generated using the COSMOS switch-level simulator on a circuit description extracted from the layout. Command files for the tests described above were created, executed by COSMOS, and the resultant output converted to Tektronix S3260 format. The functional test comprised (1) an exhaustive exercise of all circuit nodes; (2) for each of three sets of coefficients, operating with 80 data pairs of random input; and (3) convolution of a 4×4 block of maximum inputs. It took 10162 clock cycles.

3.6.3.3 Test Results

Eighteen packaged devices were received from a fabrication in 3- μ m p-well in MOSIS run M81W. Seventeen chips functioned correctly at 10 MHz; one functioned correctly but only to 1 MHz. In tests to explore high-frequency capabilities, there was full functionality to 18 MHz with $V_{dd} = 5$ V. Output sampling time at the upper limit was critical: the 10-ns sampling interval had to begin from 0 to 10 ns after the clock rising edge. Circuits functioned to 20 MHz when test-point outputs **TP1** and **TP2** were not sensed.

Output characteristics of one chip are given in Table 3-2. Delays were measured between midpoints of clock and output transitions; transition times, between the 10- and 90-percent points. Load on the output drivers from S3260 sensing circuits and the oscilloscope probe is about 50 pF, so the transition times indicate a drive capability of about 7 mA. Measured delay is the sum of delays in *filter* clock drivers, shift register output, and data driver. The discrepancy between rise and fall delay is believed to arise in the shift-register flip-flop. Note the long delays for fall transitions at both test points. **TP1** delay is more than a clock cycle at 20 MHz, which can explain functional failure at high frequency when this output is sensed. However, **TP2** delay is harmless, for output samples in testing this parameter-loading circuit are taken 2 clock cycles after a change.

TABLE 3-1.

Coefficient for Different Number of Good Inputs

n	M	
	Decimal	Binary
5	819	XX001100110011
4	1024	XX010000000000
3	1365	XX010101010101
2	2048	XX100000000000
1	4095	XX111111111111

TABLE 3-2.

Filter Output Timing

Output	Comparison Clock	Delays (ns)		Transitions (ns)	
		Rise	Fall	Rise	Fall
Sout	PHI	33	45	29	33
OUT	PHI	33	45	25	30
TP1	PHI	42	53	22	29
TP2	iclk	39	50	25	33

4. WAFER DESIGN

4.1 FLOOR PLAN

A wafer floor plan is a layout of cells and tracks on the wafer to approximate dimensions. A precise layout is not possible until design of the circuit cells is complete, but the floor plan gives guidance on desired shape of circuits and placement of I/O pins. After considering several options, it was decided to place cells of one type in columns with wafer input and output signal distribution in vertical channels and most cell-to-cell connections and global signals in horizontal channels. Therefore, the circuit cells were laid out to all have about the same height. As seen in Table 2-1, all cells except *delay* are about 2.5 mm high and *delay* is half of that. The floor plan also determined placement of pins on the circuit cells.

Figure 4-1 is a wiring diagram for the five processors of one wafer. There is no significance to the size of the cells but the placement of I/O pins on the cell sides is correct except for *delay*. Each row comprises one processor, and from left to right the cells are: five *input*, ten *delay*, and one each *threshold*, *average*, and *filter*. Global signal inputs are on the left, data inputs on the top left, and data outputs on the top right.

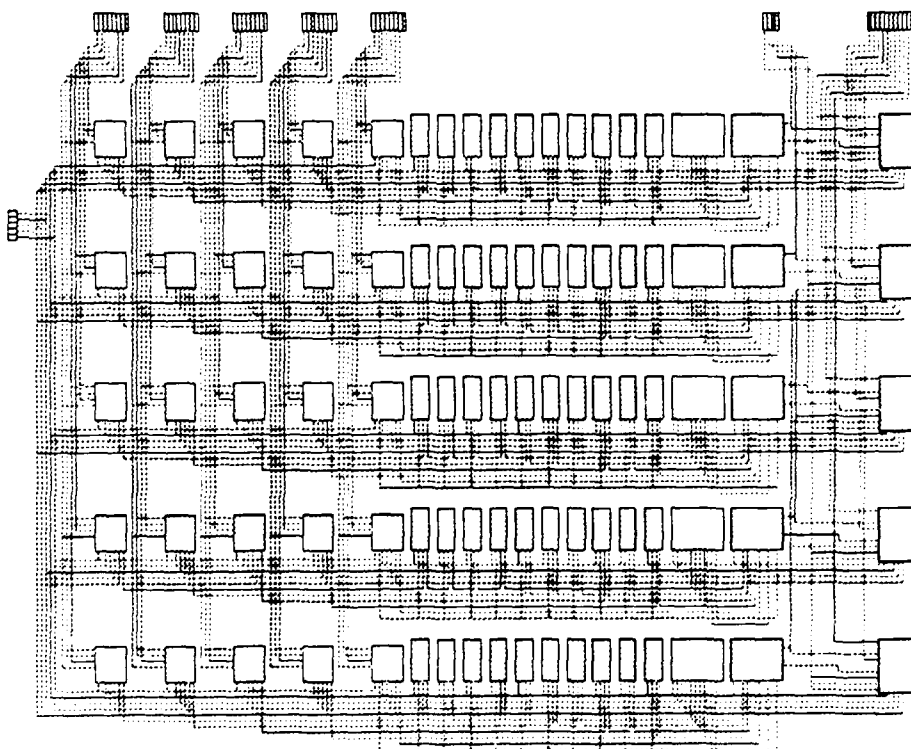


Figure 4-1. Wiring diagram of one FPP wafer.

Each wafer must have about twice as many cells as required to build the system. Fabrication on 3-in wafers would normally provide 50-mm² pieces of silicon, but with SOI wafers more edge margin is required, and 45 mm is maximum. With this constraint it is not possible to get 100 percent redundancy so 80 *delay* cells are provided for the 50 required. Since the *delay* cell size is half as large as others, the lesser redundancy should be acceptable. Figure 4-2 shows the floor plan of the top half of a wafer; the bottom half is similar. This figure was created by the Floorplanner program [7]. It is not to scale, but the circuit cells are approximately scaled relative to each other. Power buses, ~0.3 mm wide, will be in the horizontal channels and will be bonded out on the left and right sides. Each horizontal channel will have 20 signal tracks which will be placed on ~25- μ m centers. All wafer I/O circuits will be on the top and bottom of the wafer. A complete wafer layout has not been done, but it is estimated that the active area will be 44 mm wide and 41 mm high.

4.2 ROUTING EXPERIMENTS

A number of routing experiments with simulated circuit yield have shown that the number of interconnect tracks is adequate. Figure 4-3 shows a routing of all signals for one assignment of cells on the wafer. This plot is output by the IRT program which did the routing of signals.

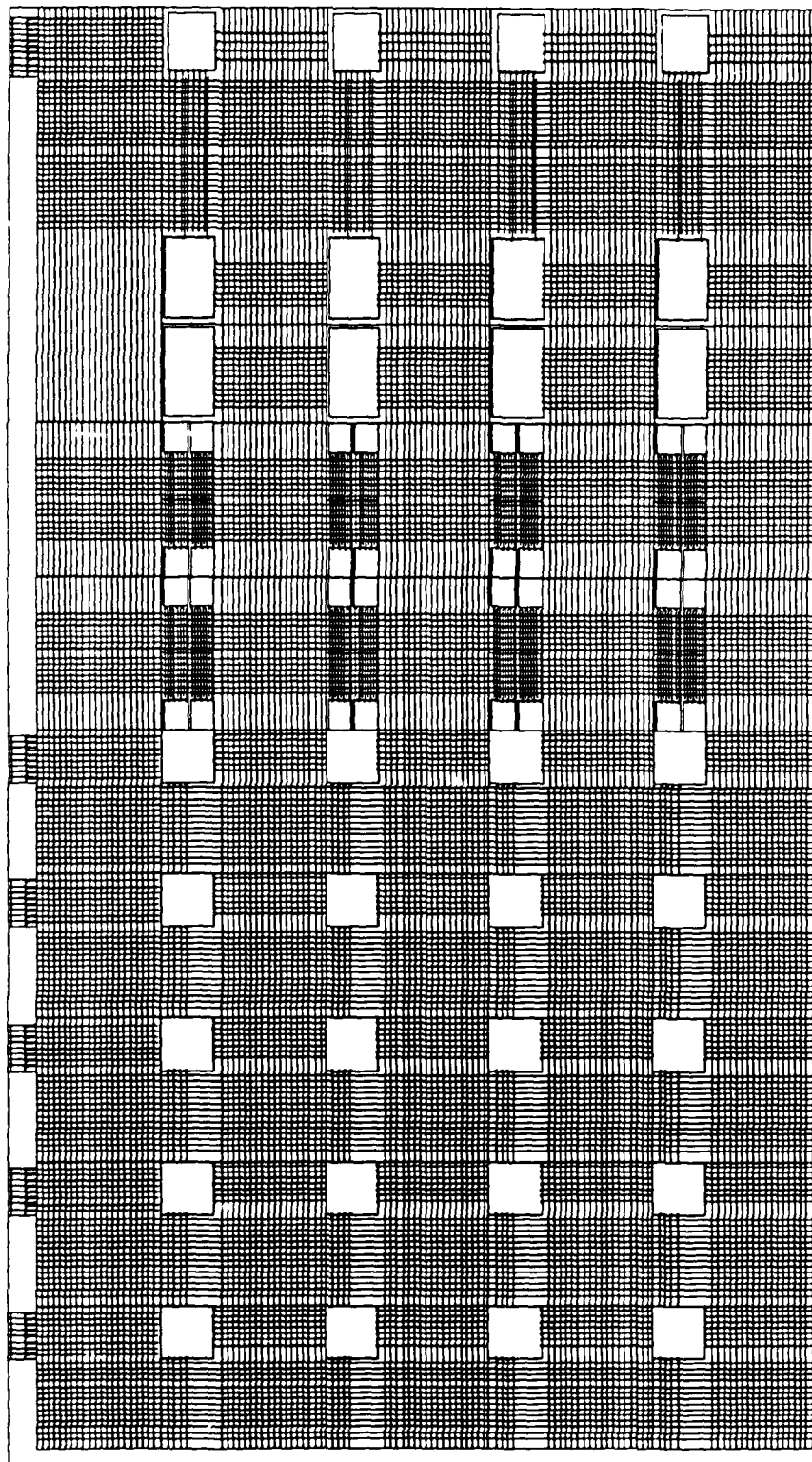


Figure 4-2. Floor plan of the top half of the FPP wafer.

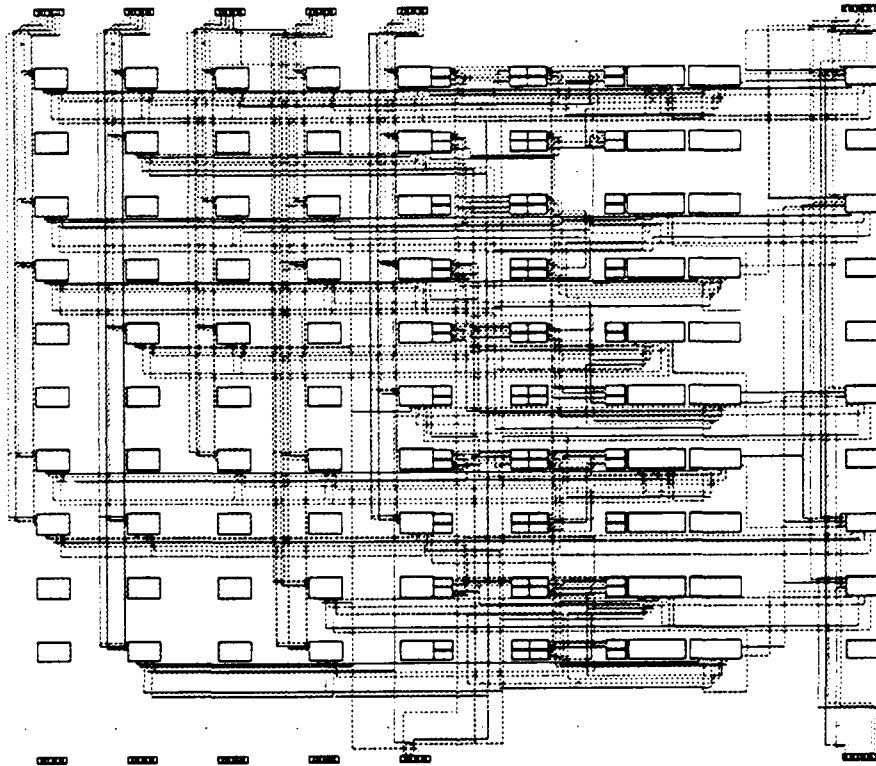


Figure 4-3. Routing of signals on an FPP wafer.

127224-8

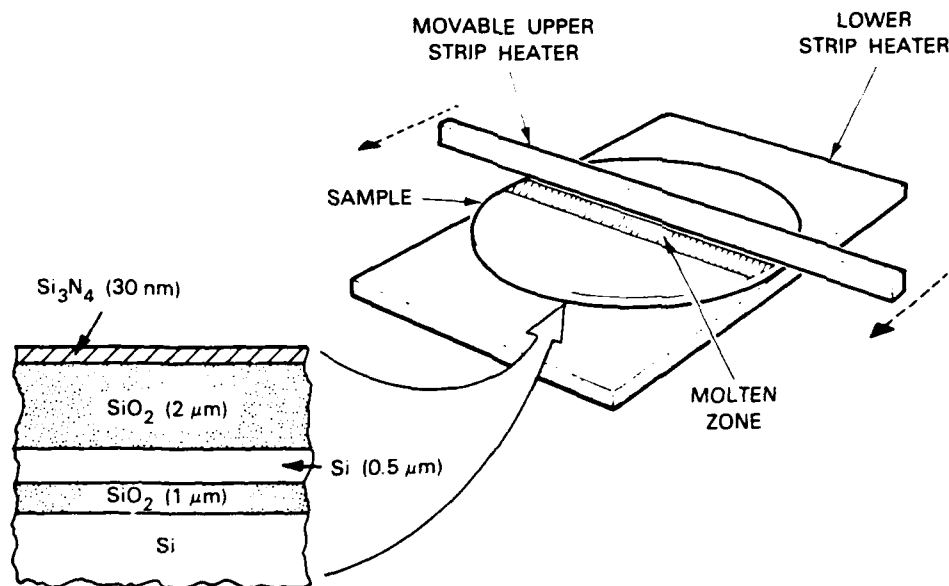
5. SOI FILMS BY ZMR

In recent years, substantial efforts have been directed toward the development of a technology for producing high-quality single-crystal semiconductor films on insulating substrates. These efforts have been motivated by the potential of thin-film devices for achieving higher packing density, speed, and radiation resistance than bulk devices, and by the potential of SOI structures for accomplishing the three-dimensional integration of electronic circuits. With support from the Radiation-Hardened Wafer Scale Program, our principal goals have been to develop a ZMR process to prepare SOI films suitable for rad-hard WS integration and to provide high-quality ZMR SOI wafers to fabricate rad-hard WS circuits. In this report, we will begin with a brief description of the problems facing ZMR SOI at the start of the program. Next, we will describe the design, fabrication, and operation of a new ZMR system, the development of a reliable capping technique, and the resulting improvements in the overall quality of recrystallized SOI films. Presented will be the results of a study correlating liquid-solid interface morphologies observed in situ during ZMR, to the defects observed in the SOI film after ZMR. Also discussed will be the role of radiative heating in ZMR and the development of a new ZMR configuration with enhanced radiative heating to prepare subboundary-free SOI. Finally, we will describe our effort to ensure a successful transfer of technology to industry and discuss the remaining materials issues in ZMR SOI.

5.1 ZMR BY THE GRAPHITE-STRIP-HEATER TECHNIQUE

The configuration of the original graphite-strip-heater system used for ZMR of Si on SiO₂ is shown schematically in Figure 5-1. The sample, which is placed on the stationary lower heater, consists of a fine-grained Si film on an insulating substrate or layer, together with an encapsulation layer over the Si film. The inset of Figure 5-1 shows a schematic cross section of a typical sample, prepared by coating a Si wafer 10 to 20 mil thick with a 1 μ m thick, thermally grown SiO₂ layer, a 0.5- μ m poly-Si layer formed by low-pressure chemical vapor deposition (LPCVD), a 2- μ m layer of CVD SiO₂, and a 30-nm layer of sputtered Si-rich Si₃N₄. The lower strip heater is used to heat the sample to a base temperature of 1100 to 1300°C, generally in a flowing Ar gas ambient at atmospheric pressure. Additional radiant energy, provided by the movable upper strip heater, is used to produce a narrow molten zone in the poly-Si layer (mp of Si = 1410°C). The molten zone is then translated at \approx 0.5 mm/s, leaving a recrystallized Si film. The thicknesses of the SiO₂ and Si layers, the composition of the encapsulation layer, the molten zone speed, and the shape of the upper heater and its position relative to the sample, all have a strong effect on the morphology and crystallography of the recrystallized films.

Although the original ZMR system produced a good yield of device-quality material, it did not yield uniform films routinely and could not be used to process wafers of >3-in diam., seriously limiting collaboration with other laboratories. The principal material defects in the SOI films were low-angle grain boundaries (subboundaries). Lack of good thermal uniformity resulted in significant variation in the material quality over the film surface and wafer warpage unacceptably high for standard VLSI processing. Substrate melting occurring near the edge of the SOI wafers during ZMR seriously reduced the useful recrystallized area for device fabrication. Lack of mechanical stability



127224-9

Figure 5-1. Schematic diagram of graphite-strip-heater system used for ZMR of encapsulated Si films. The inset shows a cross section through a typical sample.

resulted in poor run-to-run reproducibility, and the effectiveness of the Si_3N_4 – SiO_2 encapsulation layers in preventing agglomeration of the SOI film during ZMR was unpredictable, making it very difficult to conduct controlled experiments.

5.2 NEW SYSTEM FOR ZMR

To overcome the limitations of the original system, a new strip-heater system was constructed incorporating many improved features and, when suitably fixtured, capable of processing 6-in wafers. This new system has permitted the reproducible preparation of uniformly recrystallized 4-in films and also has led to a significant improvement in the quality of films which are $\leq 0.5 \mu\text{m}$ thick. Figures 5-2 and 5-3 are photographs of the new system's exterior and interior, respectively.

With the objective of routinely achieving uniform edge-to-edge recrystallization, we adopted the following design goals for the new ZMR system: uniform base temperature over the entire surface of the Si film; constant spacing between the upper strip heater and the wafer surface throughout the heater scan; smooth motion of the liquid-solid interface for scan speeds from 25 $\mu\text{m/s}$ to 2.5 mm/s; and the mechanical, chemical, and thermal stability required for run-to-run reproducibility.

78872-19

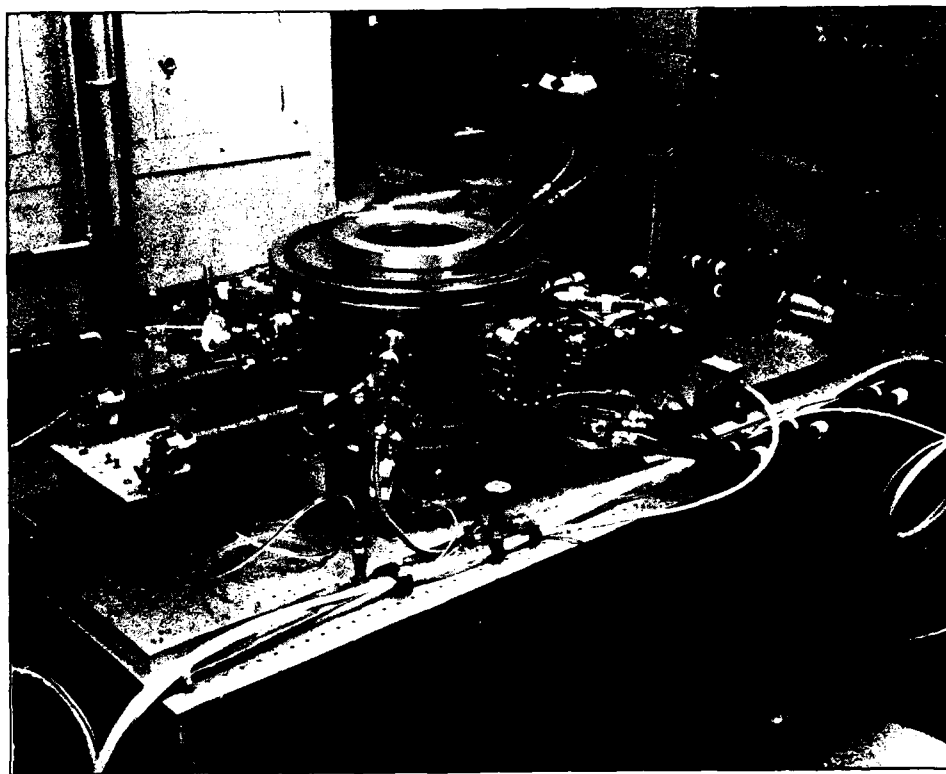


Figure 5-2. External view of new system for ZMR of SOI films.

78872-20

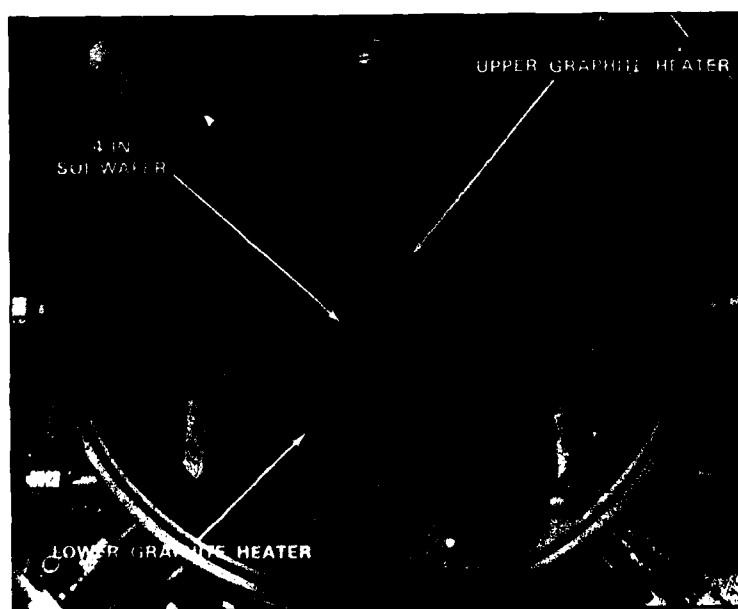


Figure 5-3. Top view of interior of new ZMR system.

To obtain a uniform base temperature, the wafer is placed on an independently mounted graphite platen rather than directly on the lower heater. The platen, which increases the thermal mass and reduces the effect of radiation from the upper heater on the base temperature, has tantalum heat shields, an auxiliary tungsten-wire ring heater, and a series of nested graphite inserts to minimize radial heat flow. The platen is mounted by means of machined quartz fixtures that provide thermally stable mounting with minimal conductive heat loss. Heat loss from the lower heater to the chamber is reduced by using a series of graphite and tantalum heat shields at the edges and bottom of the heater.

In order to maintain a fixed gap between the upper heater and the wafer surface, the upper and lower heaters are preloaded using tantalum-tungsten alloy springs to accommodate thermal expansion and prevent bowing of the heaters. The platen and upper heater can be positioned independently in order to provide the degrees of freedom necessary to keep the gap constant during scanning.

To ensure that the solid-liquid interface moves smoothly during ZMR, the new system is designed to minimize jitter in the motion of the upper heater and vibration arising from other sources. The heater strip is clamped at each end to a quartz rod extending in the direction of zone motion. The opposite ends of each quartz rod pass through openings in the chamber wall and into flexible metal bellows that are sealed to the wall by flanged metal tubes. Thus, there are two bellows on each side of the chamber. Each bellows is sealed at the far end by a flanged metal plate, and the end of the quartz rod is clamped to this plate. Each pair of flanged plates is mounted on a metal plate attached to a linear air-bearing slide. On one side of the chamber this mounting plate is attached to a high-precision lead screw that is driven by a dc servomotor. By using the air-bearing slides and the bellows, which expand and contract with the motion of the lead screw, this motion is transmitted smoothly to the two quartz rods and thus to the upper heater, while the vacuum integrity of the chamber is maintained without the use of sliding seals or rotary feedthroughs. Several features are incorporated in order to minimize vibration due to sources other than the heater motion: the chamber and drive mechanism are mounted on an isolation table; the turbomolecular pumping system used to evacuate the chamber is mounted directly below the chamber and vibrationally isolated by means of a damped bellows assembly; cooling-water turbulence is reduced by using several parallel water lines with large-bend radii for cooling the chamber and the large O-ring seals between the top cover and chamber body, and also between the top cover and quartz viewing window.

The ZMR chamber is basically an ultrahigh-vacuum chamber which provides a clean processing environment and permits the use of commercially available components, such as viewports, electrical feedthroughs, and vacuum valves. For long-term stability of the system, all components that are heated to about 500°C or higher during ZMR are fabricated of high-temperature materials, including quartz, boron nitride, alumina, graphite, tantalum, and tungsten. The only other materials used in the chamber are copper and stainless steel, which are thermally shielded, water cooled, or heatsunk to prevent them from heating to excessive temperatures.

To ensure run-to-run reproducibility, the new ZMR system is operated with the aid of an IBM PC. A computer interface was installed to permit automated programmable control of the

lower and movable upper graphite strip heaters, the auxiliary tungsten ring heater, and the speed and direction of motion of the upper heater. As illustrated in Figure 5-4, the computer can be

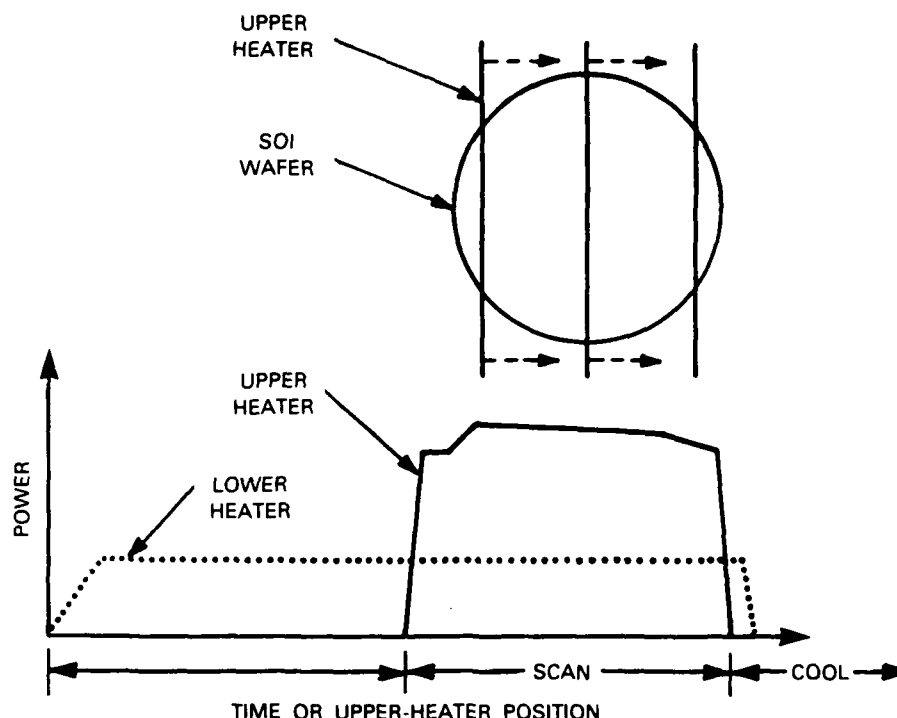


Figure 5-4. A schematic illustration of computer control of the lower and movable upper heaters.

used to control the power of the heating elements in order to accommodate slight changes in the base temperature that occur during ZMR, ensuring uniform recrystallization of the SOI film over the entire surface and preventing overheating of the film near the beginning and end of the scan. By supporting the SOI wafer around the perimeter with an edge heatsink, we further improved the temperature uniformity of the film during ZMR and successfully suppressed substrate melting originating near the wafer edge. Using this technique, 3- and 4-in wafers have been recrystallized to within 3 mm of the perimeter, free of edge-related macroscopic defects. With the improved thermal uniformity and computer control of the new system, the warpage of a typical recrystallized 3-in SOI wafer has been reduced to less than 40 μm .

5.3 NEW CAPPING TECHNIQUE FOR ZMR

In the ZMR process for preparing SOI films, the poly-Si film on the SiO₂-coated Si substrate is recrystallized by the passage of a molten Si zone. To obtain a device-quality SOI film, the poly-Si film must be encapsulated to ensure that the molten Si zone will exhibit uniform wetting as it traverses the film. In this section, we discuss a new capping technique that has made it possible to achieve a major improvement in the effectiveness and reproducibility of ZMR performed by the graphite-strip-heater technique.

In an earlier study, we obtained substantial evidence that SiN_x encapsulation promotes wetting because a trace amount of N diffuses through the capping SiO₂ layer and is incorporated at the CVD-SiO₂ /poly-Si interface. In earlier work on samples prepared by growing a film of thermal SiO₂ \approx 400 Å thick on a Si wafer, it was shown that high-temperature annealing in NH₃ caused the introduction of N into the SiO₂ film, and that the N concentrations were higher at the upper and lower boundaries of the film than in the interior. This is the same kind of process used for nitridation of gate oxide, as described in Chapter 7. These findings suggested that NH₃ annealing of SOI samples encapsulated with SiO₂ might cause incorporation of sufficient N at the upper SiO₂/Si interface to promote wetting by the molten Si during ZMR. This was found to be the case. For samples with the geometry shown in Figure 5-5, high-temperature annealing at 1100°C for 3 h in NH₃, oxidizing for \approx 20 min in O₂, and annealing in NH₃ for an additional 3 h, consistently results in excellent wetting and <100> texture of the SOI film. Before each annealing step the system is purged with N₂ to prevent reaction between NH₃ and O₂. Annealing in NH₃ for the same total time without an intermediate oxidation step is less effective. Exposure to NH₃ produces a thin oxynitride film on the SiO₂ surface that apparently impedes incorporation of N into the SiO₂ and removal of this layer by oxidation permits incorporation to proceed more rapidly.

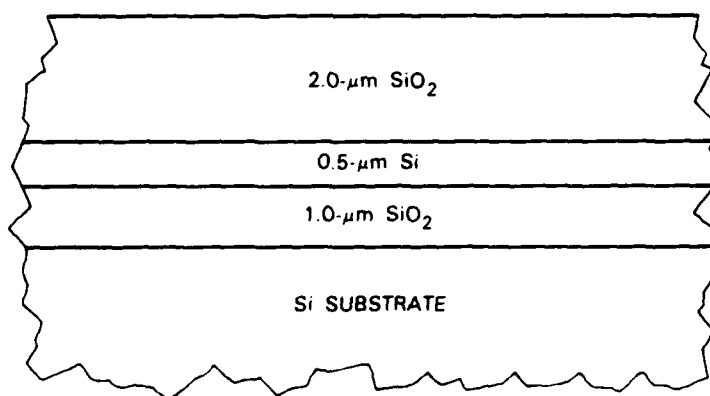


Figure 5-5. Schematic cross section of a typical SOI wafer.

Auger electron spectroscopy was used to investigate the N concentrations that are introduced by NH₃ annealing. From measurements on a control sample prepared by the deposition of Si₃N₄, we estimate that our detection limit for N at the SiO₂/Si interface is approximately half a monolayer. For a sample with the configuration of Figure 5-5, but with an SiO₂ capping layer only 0.2 μm thick,

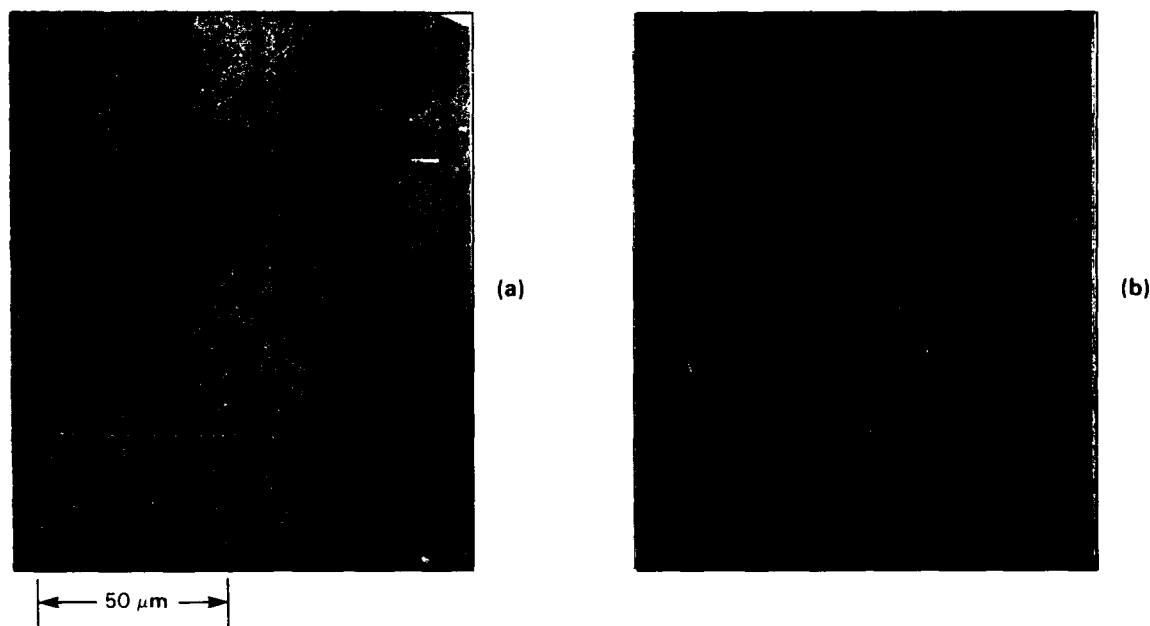


Figure 5-6. Two optical micrographs of the same area of a recrystallized SOI film after defect etching, illuminated by (a) monochromatic radiation and (b) white light. A grain boundary is seen on the left, showing that the defect etch was effective.

N was detected at the SiO_2/Si interface after annealing in NH_3 at 1100°C for 1 h. For a sample with a standard $2\text{-}\mu\text{m}$ SiO_2 capping layer, after $\text{NH}_3/\text{O}_2/\text{NH}_3$ annealing sufficient to produce excellent wetting during ZMR, no N was detected by Auger analysis at the SiO_2/Si interface. We estimate that the amount of N present at the interface after this annealing treatment is roughly one-third of a monolayer, in view of the results on the sample with the thinner SiO_2 layer.

The NH_3 annealing process has several important advantages over SiN_x capping. Annealing, which is performed in a standard diffusion furnace, is much less susceptible to contamination, especially from particulates, than sputtering. The N concentration incorporated at the SiO_2 interface can be accurately adjusted by fixing the annealing time, temperature, and NH_3 partial pressure. In addition, the interface concentration is uniform over the entire wafer because N diffusion occurs under controlled conditions during the annealing process; SiN_x capping can yield inhomogeneous N distributions because diffusion occurs in the graphite-strip-heater system during the heatup period just before recrystallization. Because of these features of the NH_3 process, ZMR using this process reproducibly yields SOI films 0.3 to $0.5\text{ }\mu\text{m}$ thick that are extremely smooth and uniform in thickness.

The SOI films prepared in earlier ZMR experiments using SiN_x capping generally contain a high density of branched subboundaries. In recent experiments using such capping, however, $1\text{-}\mu\text{m}$ SOI films scanned at 0.5 mm/s were found to contain large areas with unbranched subboundaries as well as some regions with only trails of dislocation clusters and diffuse bands of dislocations. We have obtained still better results in similar experiments using a single 8-h NH_3 anneal of the SOI wafer before ZMR, which introduces less N than the $\text{NH}_3/\text{O}_2/\text{NH}_3$ treatment described above. Figure 5-6 shows optical micrographs, taken after Secco etching for defect delineation, of a portion

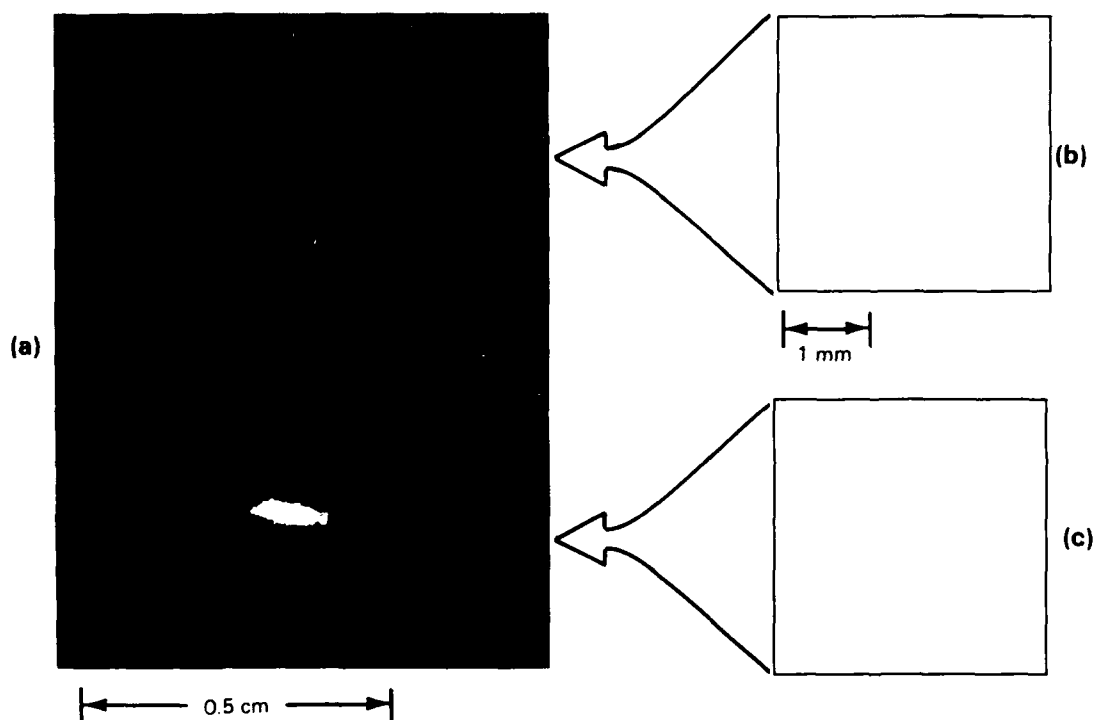


Figure 5-7. (a) Optical micrograph of the SOI film of Figure 5-6. The scan direction was from top to bottom. (b), (c) Electron channeling patterns obtained by operating a scanning electron microscope in the backscattering mode. The dashed line in (b) shows the location of a grain boundary.

of a recrystallized $1\text{-}\mu\text{m}$ SOI film with a $2\text{-}\mu\text{m}$ underlying SiO_2 layer, scanned at 0.5 mm/s . Very few subboundaries are observed, although small ridges or slight thickness variations occur where subboundaries might be expected. The only defects observed over ≈ 80 percent of a 3-in wafer (other than those associated with grain boundaries like the one seen at the left of Figure 5-6, are trails of isolated dislocations, which have been shown by transmission electron microscopy to be threading dislocations running nearly normal to the surface. The density of these defects averaged over an area of several square centimeters is typically less than $2 \cdot 10^6\text{cm}^{-2}$.



Figure 5-8. Optical micrograph of the low defect density region of Figure 5-7(a) at higher magnification.

The critical influence of the experimental conditions on the quality of SOI films is illustrated by Figures 5-7 and 5-8. Figure 5-7(a) is an optical micrograph of a larger region of the recrystallized film shown in Figure 5-6. Except for several grain boundaries, which can easily be prevented by seeding to the underlying substrate, isolated threading dislocations are the only defects present in the upper portion of the sample. The lower portion shown in Figure 5-7(a), which was recrystallized after an increase of ≈ 5 percent in the power to the upper-strip heater, contains a high density of subboundaries. Figures 5-7(b) and 5-7(c) are scanning electron micrographs of the two portions of the sample taken in the backscattering mode. A very distinct channeling pattern, indicating a high degree of crystalline perfection, is observed for the upper portion, while the pattern for the lower portion clearly shows discontinuities associated with the subboundaries.

The results shown in Figure 5-7 support the hypothesis that stresses produced by thermal gradients in the substrate are responsible for the formation of subboundaries. According to this hypothesis, subboundaries were not formed in the upper portion of the film because the thermal gradients at the liquid-solid interface and the associated stresses in the Si film and substrate were relatively low. The increase in power to the upper heater during recrystallization of the lower portion increased the thermal gradients and therefore the stress in the film. The increase in stress resulted in the plastic deformation of the film and the formation of subboundaries.

Figure 5-8 is a higher magnification optical micrograph of the area of low defect density shown in Figure 5-7(a). The dislocations in such areas are often associated with needle-like protrusions, one of which is seen in the lower right corner of Figure 5-8. These protrusions, which are believed

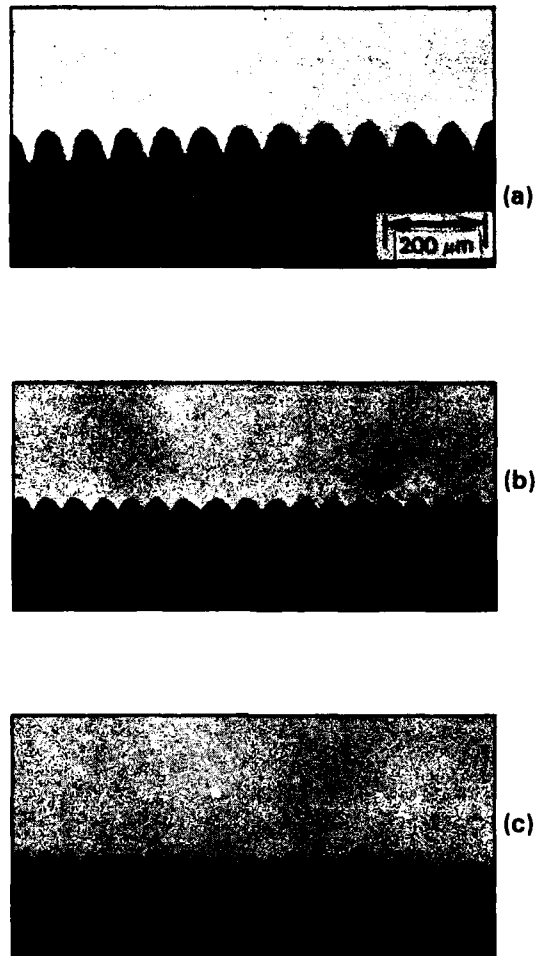


Figure 5-9. Interface morphologies as a function of increasing upper-strip power at (a) 2.13, (b) 2.49, and (c) 2.54 kW.

to result from the presence of impurities, have been shown by TEM to be crystallographically continuous with the adjacent Si film. Optical microscopy shows that they point predominantly in the ZMR scan direction.

5.4 LIQUID-SOLID INTERFACE MORPHOLOGIES DURING ZMR

With the ZMR process now capable of producing high-quality SOI films in which isolated dislocations or dislocation clusters are the principal defects, we conducted a study to define the

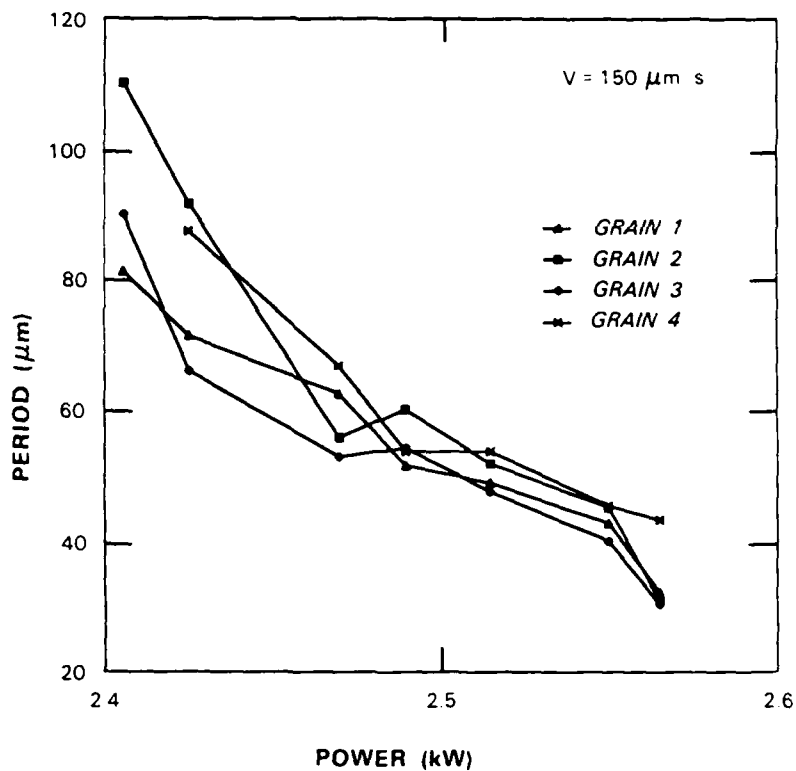


Figure 5-10. Cell spacing for four grains as a function of upper-strip power.

experimental conditions that yield the lowest defect densities. To prepare the samples, a 1- μm -thick amorphous Si film was formed by low-temperature low-pressure chemical vapor deposition (LTLPCVD) on (100) Si wafers coated with 2 μm of thermally grown SiO_2 . The Si film was capped with a 2- μm -thick LPCVD film of SiO_2 , and the samples were annealed at 1000°C in ammonia. In-situ observations of the solidification interface were then made during ZMR in a graphite-strip-heater system equipped with an optical microscope and a video recorder. In this section the relation between the observed liquid-solid interface morphologies and the corresponding defect structures is described.

Two sets of experiments were performed. In each run of the first set, the power to the upper-strip heater was varied in steps between about 2.3 and 2.6 kW, with the zone velocity kept at 150 $\mu\text{m}/\text{s}$. In each run of the second set, the velocity was varied in steps from about 100 to 450 $\mu\text{m}/\text{s}$ with the upper-strip power kept constant at ≈ 2.5 kW. In all cases the separation between the sample and upper heater was 0.8 mm. Optical micrographs of the liquid-solid interface taken during ZMR at 150 $\mu\text{m}/\text{s}$ are shown in Figure 5-9. As the power to the upper strip was initially increased, partial melting of the Si film occurred. When the power was increased sufficiently to produce a completely molten zone, a cellular-dendritic liquid-solid interface morphology was observed. As the power was further increased, a simple cellular morphology was obtained [Figure 5-9(a)], and there was a gradual decrease in both the cell period and the amplitude of the interface structure

[Figures 5-9(b) and (c)]. The change in cell period is illustrated in Figure 5-10 by results for four different grains. When the power was increased above about 2.6 kW, the interface developed facets of very small amplitude.

In the cellular morphology regime, post-solidification etching and optical microscopy revealed three types of defects in the recrystallized films: isolated dislocations, dislocation clusters, and subboundaries (Figure 5-11), which were dominant at low, intermediate, and high upper-strip power levels, respectively (Figure 5-12). It has been demonstrated that the line directions of the isolated and clustered dislocations are perpendicular to the plane of the film, while subboundaries are low-angle grain boundaries composed of dislocations that lie both perpendicular and parallel to the plane of the film.

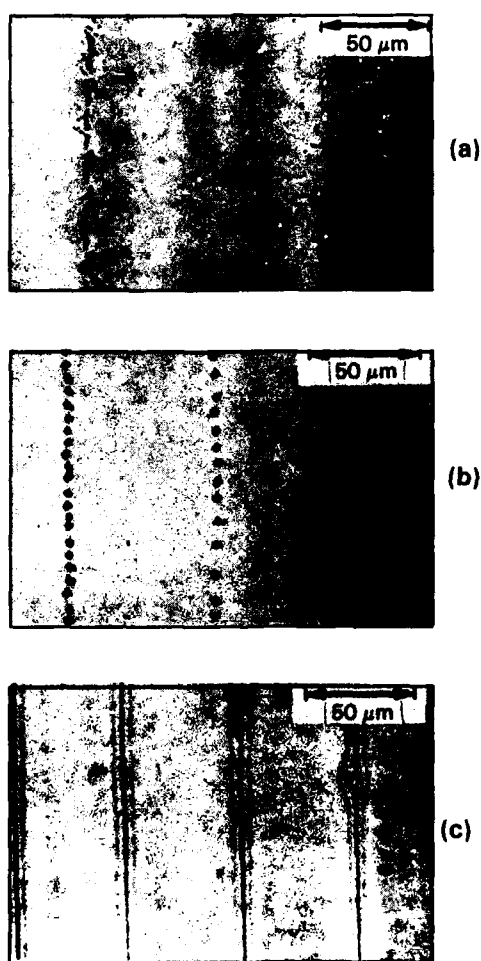


Figure 5-11. Optical micrographs showing three types of defects, (a) trails of isolated dislocations, (b) trails of dislocation clusters, and (c) continuous subboundaries.

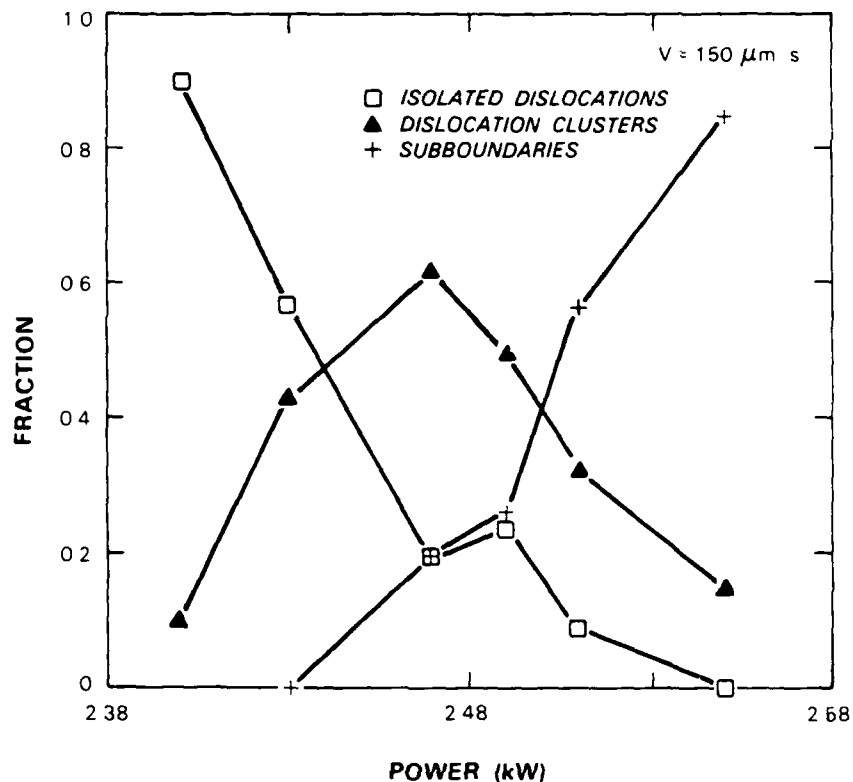


Figure 5-12. Fraction of the three defect types shown in Figure 5-11, plotted as a function of power. (Data were taken from a single grain.)

Whereas cell period and amplitude both decreased monotonically with increasing upper-strip power, variations in the velocity at a constant 2.5 kW led to more complex changes in interface morphology (Figure 5-13). For low velocities, increasing velocity produced a gradual increase in the cell period (Figure 5-14) and the depth of the interface cusps also increased. Following each change in velocity, a steady state was reached in a few seconds, after which new cells were not created nor existing cells annihilated. This stability resulted in the formation of parallel defect trails in the zone-melting films. When the velocity was increased above about 300 $\mu\text{m/s}$, the cell structure became unstable. Cells would typically split in the middle [Figure 5-13(c)], and new cusps would continuously develop and vanish. With further increases in velocity, the period of this unstable interface decreased (Figure 5-14).

The changes in velocity also led to changes in the types of defects formed. For low velocities, isolated dislocations and dislocation clusters were observed [Figure 5-15(a)]. For higher velocities, in addition, X-shaped dislocation clusters [Figure 5-5(b)] were found around protrusions that presumably formed from trapped liquid droplets. When the interface became unstable, branched defect patterns were obtained [Figure 5-15(c)].

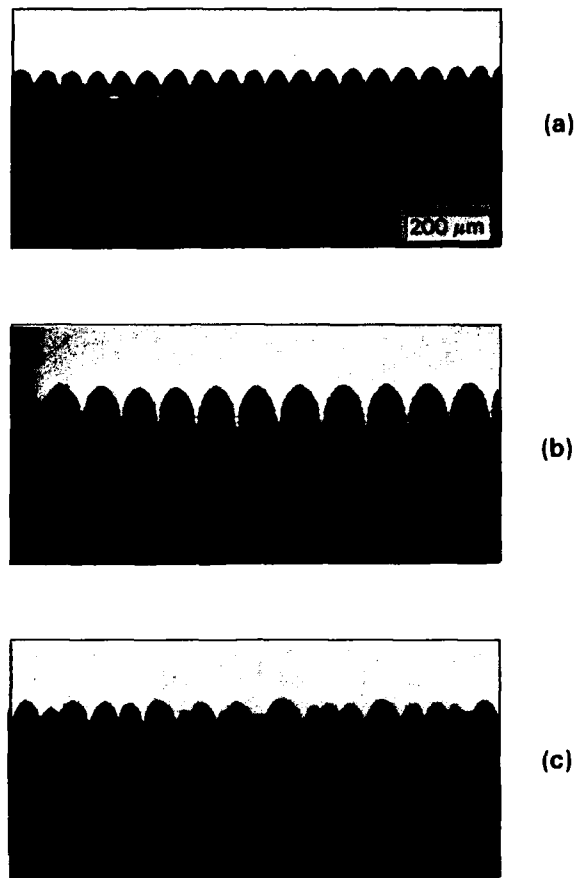


Figure 5-13. Interface morphologies as a function of increasing velocity at (a) 90, (b) 330, and (c) 420 $\mu\text{m/s}$. Upper-strip power = 2.5 kW.

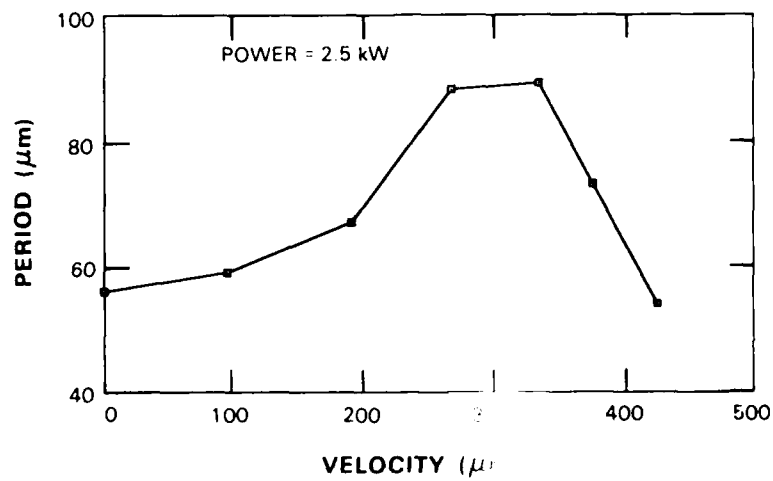


Figure 5-14. Dependence of cell period on zone velocity.

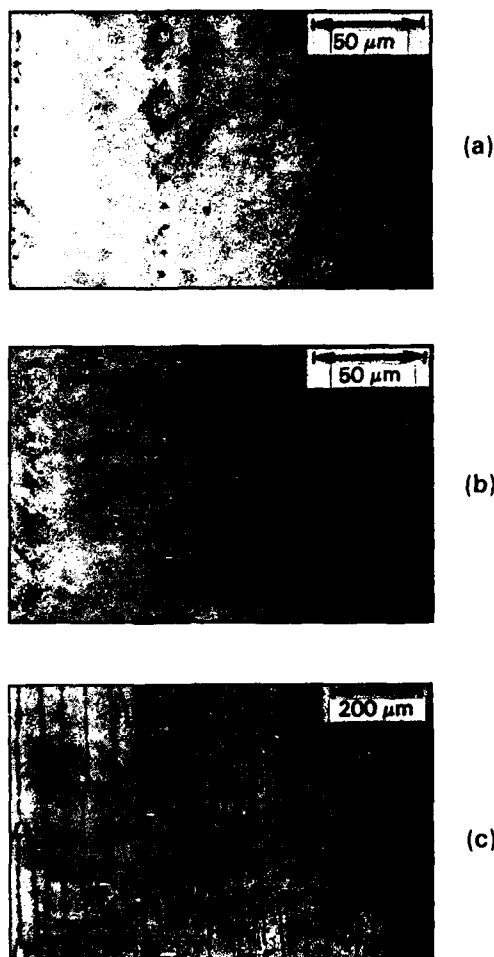


Figure 5-15. Optical micrographs showing defect patterns obtained for velocities of (a) 90, (b) 330, and (c) 420 $\mu\text{m/s}$.

Films with the lowest defect densities were obtained at low zone velocities for upper-strip power levels at the low end of the range that yields a stable cellular interface morphology. The lower limit to this power range is imposed by development of a cellular-dendritic morphology, which leads to side branches and protrusions on the surface of the recrystallized films.

We propose that radiant heating from the upper strip, rather than constitutional supercooling, accounts for the stable cellular interface morphology that is observed in the low-power, low-velocity regime yielding the lowest defect densities. If this morphology were produced by constitutional supercooling, the cell period would decrease with increasing velocity, but the data of Figure 5-14 show that in this regime the cell period increases with zone velocity.

The observed variation in cell amplitude and period with upper-strip power can be explained by changes in the radiation intensity gradient at the solidification interface, which increases with increasing power. In the low-velocity regime, the reduction in cell amplitude with an increase in power reflects the corresponding increase in the intensity gradient. The decrease in the cell period with increasing power may be qualitatively understood in terms of the amplitude-to-period ratio, which roughly reflects the radius of curvature of the cell tip. For a given period, when the cell amplitude decreases the radius of the cell tip increases. Because of the increased fraction of solid Si, this change could result in significant superheating of the solid and undercooling of the liquid at the tip, leading to an unstable interface and perturbation growth. To restore the preferred tip curvature, which depends on the steady-state undercooling and the liquid-solid interfacial energy, an additional cell would then be formed, decreasing the period.

The locations of all defects observed in the solidified films correspond to the trailing cusps between adjacent cells at the liquid-solid interface. Whatever the cause of the cellular morphology, most impurities will be rejected to the cell boundaries. The dislocations present in subboundary-free films may form, at least in part, as a result of impurity incorporation along the cell boundaries. When radiation intensity gradients are high enough, subboundary formation may result from thermal stress due to nonlinear thermal gradients adjacent to the interface region.

5.5 ZMR WITH ENHANCED RADIATIVE HEATING

When the ZMR process was first introduced, subboundaries were found to be the principal defects in the recrystallized SOI films. Subsequently, a number of groups have identified ZMR conditions under which subboundary-free 1- μm -thick SOI films can be prepared. However, it has been very difficult to eliminate subboundaries from films with a thickness of 0.5 μm or less, which are of greater importance for large- and very-large-scale integrated circuit applications. In this section we describe a novel ZMR configuration, with enhanced radiative heating, that has enabled us to prepare subboundary-free 0.5- μm -thick SOI films over a much wider range of experimental parameters than previously possible.

Schematic cross-sectional diagrams comparing the conventional and new ZMR configuration with enhanced radiative heating are shown in Figure 5-16. In the conventional configuration [Figure 5-16(a)], the sample is mounted between the upper and lower heaters, with the SOI film facing upward. In the new configuration, the sample is positioned above the movable heater with the film facing downward [Figure 5-16(b)].

Figure 5-17 has optical and transmission electron micrographs of a defect-etched subboundary-free 0.5- μm -thick SOI film prepared by ZMR using the enhanced radiative heating configuration with a tungsten upper heater 2.5 mm wide and 0.15 mm thick. The principal defects are isolated threading dislocations with a density of $\approx 10^6 \text{ cm}^{-2}$, the same level obtained in subboundary-free 1- μm -thick films. The dominant features that appear in Figure 5-17(a) are etch pits associated with these dislocations, occasional voids resulting from undercutting of the underlying SiO_2 occurring during the defect etching of the SOI film, and ridges formed along the last-to-freeze regions of the

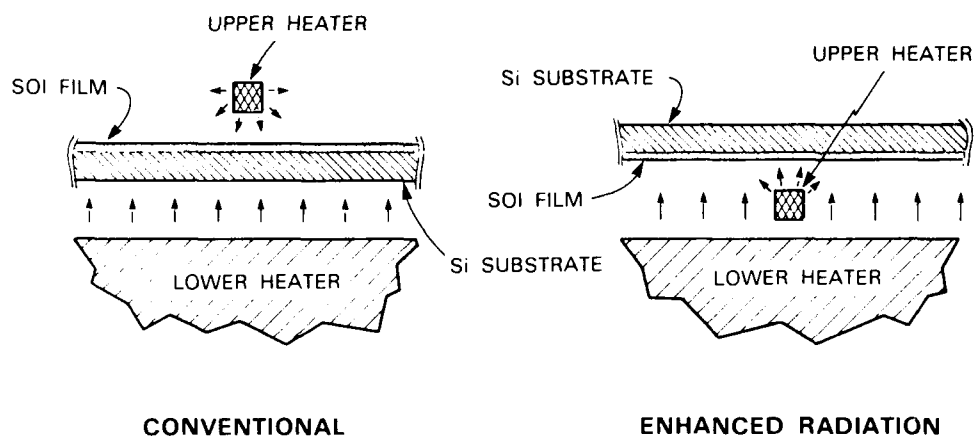


Figure 5-16. Cross-sectional diagrams showing the conventional and new ZMR configuration with enhanced radiative heating.

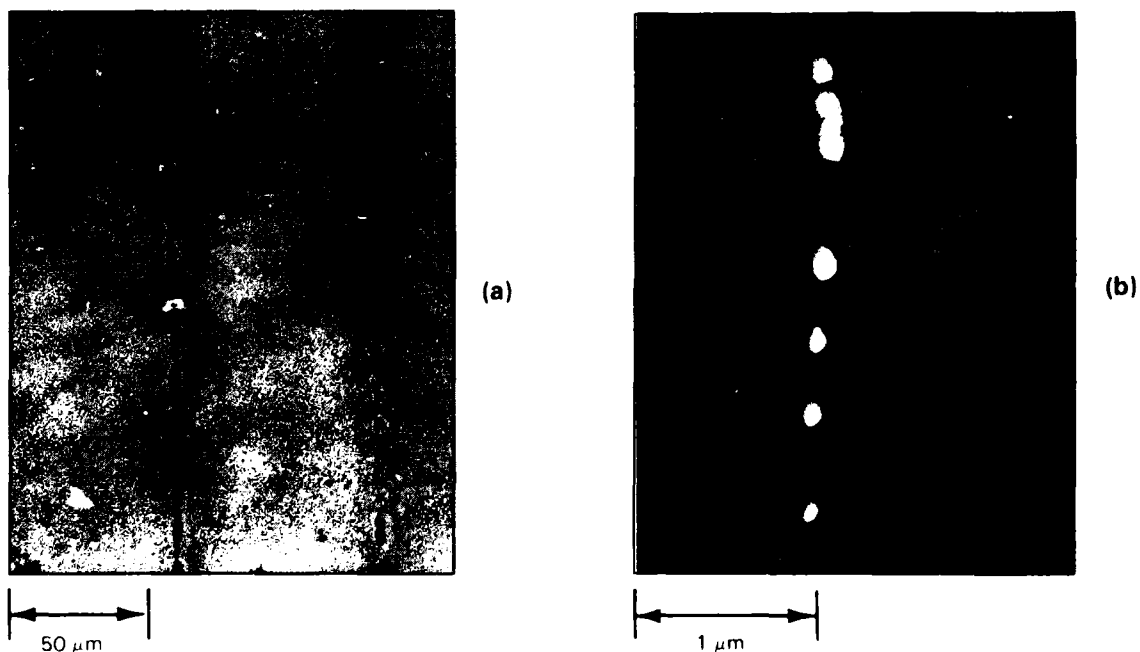


Figure 5-17. (a) Optical and (b) transmission electron micrographs of subboundary-free defect-etched 0.5-μm-thick SOI film prepared by ZMR using enhanced radiative heating.

film. The bright spots in Figure 5-17(b) are the dislocation etch pits, and the dark band is one of the ridges. Diffraction-mode transmission electron microscopy shows that the crystallographic misorientation across the ridges is less than 0.05° , the limit of measurement. The misorientation across unbranched subboundaries is typically about 0.5° . At a scan speed of ≈ 0.1 mm/s, subboundary-free $0.5\text{-}\mu\text{m}$ -thick films were obtained over a 5 percent range in movable heater power. In contrast, when the conventional configuration was used with the same scan speed, subboundary-free $1\text{-}\mu\text{m}$ -thick films could be obtained only within a range of ≈ 2 percent in movable heater power, and subboundary-free $0.5\text{-}\mu\text{m}$ -thick films were not obtained.

In previous studies using a long-working-distance zoom lens for in situ observation of the solid-liquid interface morphology, we found that subboundary-free films are obtained when the interface has a stable cellular interface with trailing cusps, provided that the heater power and scan speed are sufficiently low. The stable cellular morphology persists even when the scanning motion is stopped, showing that impurity redistribution is not the only factor responsible for this morphology. In addition, at slightly lower movable heater power levels, solid and liquid Si are found to coexist within the molten zone. This result can be attributed to the difference in reflectivity between the solid and liquid. It can be concluded that radiative heating plays a major role in determining the shape of the solid-liquid interface under ZMR conditions that yield subboundary-free films.

We believe that the greater difficulty encountered with the conventional ZMR configuration in obtaining subboundary-free $0.5\text{-}\mu\text{m}$ -thick SOI films than in obtaining subboundary-free $1\text{-}\mu\text{m}$ -thick films is explained by the fact that the thinner films absorb less of the incident radiation than the thicker ones, thereby reducing the effect of radiative heating. The relative ease of preparing subboundary-free $0.5\text{-}\mu\text{m}$ -thick films with the new configuration can be attributed principally to an increase in the radiative heating of the crystallizing interface by the movable heater. First, for the same base temperature additional movable heater power is required to melt the SOI film in the new configuration, because the back of the substrate is no longer heated by the stationary heater. Second, there is a change in the dependence of the molten zone width on movable heater power. In the conventional configuration, the stationary heater heats the SOI film via thermal conduction through the substrate, and the movable heater heats the crystallizing interface both via direct radiation and lateral conduction through the radiatively heated molten Si zone. In this case, a small increase in the movable heater power causes a large increase in the molten zone width. Therefore, the radiative heating at the solid-liquid interface is reduced, because the radiation incident on the interface rapidly decreases with increasing distance between the interface and the upper heater. This reduction in radiant heating and increase in lateral conduction leads to a more nearly planar interface, which results in closely spaced subboundaries in the recrystallized SOI film. In the new configuration, the back of the substrate is strongly cooled by radiation, enhancing the cooling of the film via thermal conduction through the substrate. Consequently, an increase in the movable heater power leads to a much smaller increase in the molten zone width and therefore to a much smaller decrease in the radiative heating of the crystallizing interface by the upper heater.

In the new configuration, the radiant heating of the SOI film is increased not only by movable heater effects described above, but also because the film is now heated by direct radiation from the stationary heater. The stationary heater makes a relatively small contribution to the radiant

heating, however, because its temperature is much lower than that of the movable heater, so that its emission intensity is lower and its radiation spectrum has less overlap with the absorption spectrum of the solid Si at the melting point.

It is encouraging that the initial experiments with the new ZMR configuration have yielded such promising results. Many key experimental parameters can be optimized to improve the crystalline quality of SOI films prepared by using this configuration. The radiation intensity profile incident at the crystallizing interface can be adjusted by varying the base temperature of the substrate, changing the shape of the movable heater and the gap between this heater and the sample, and heating or additional cooling of the back of the substrate. Such changes could result in improved surface morphology and reduced wafer warpage. In addition, the radiative spectral output of both the stationary and movable heaters can be shifted by using materials with different emissivities and spectral characteristics.

5.6 TECHNOLOGY TRANSFER AND REMAINING MATERIALS ISSUES

Under the auspices of the MIT Technology Licensing Office, we have transferred the ZMR technology developed at Lincoln Laboratory to Kopin Corporation for commercialization. Details of the design, fabrication, and operation of the new ZMR system were provided. Sample preparation procedures, including details of the new capping technology critical to preparing high-quality ZMR SOI, were also disclosed. We have continued to provide Kopin with new information as our understanding of the ZMR process has improved. The Kopin system, based upon the Lincoln Laboratory design, includes automated wafer handling, cassette-to-cassette operation, and 6-in-diam. wafer capability, with potential throughput of ≈ 10 wafers/h. ZMR SOI wafers produced by Kopin have been commercially available since early 1988.

Although the Radiation-Hardened Wafer Scale VLSI Program has been terminated, a number of materials issues for preparing ZMR SOI remain. Research to identify the primary cause of the residual defects and to model the heat flow during ZMR would permit substantial progress to be made toward the elimination of substrate slip and further reductions in thickness variation and defect densities in SOI films 0.5 to 1 μm thick. Recent studies have predicted that considerable improvements in performance could be achieved for devices fabricated in very thin films, that is ≈ 500 Å thick. Additional work is necessary to develop a ZMR process to prepare such thin SOI films free of subboundaries, with low defect densities, minimal thickness variation, and smooth surface morphology.

6. DEVELOPMENT OF A WAFER SCALE IC PROCESS FOR SOI SUBSTRATES

6.1 BACKGROUND

Ionizing radiation incident on the pn junctions of semiconductor circuits produces photocurrents which interfere with operation of the circuit. These currents can be so high that in addition to information loss and operational upset, conductors can actually be burned out. By building the circuit in a thin film of silicon on an insulating substrate, the volume of pn junctions can be reduced by a large factor, with a corresponding reduction in the photocurrent. This transient radiation effect is the motivation for use of SOI in the FPP project. The goal for the first generation FPP was to develop a process suitable for RVLSI on SOI with 3- μ m design rules, transient radiation hardness consistent with SDI Level I goals, and total dose hardness of 1 Mrad(Si).

ZMR (discussed in Section 5) was developed at Lincoln Laboratory with quality such that 1 k-bit static RAM chips and 1.2 k-gate gate array circuits had been produced. New ZMR fabrication equipment was being installed which would increase the quantity and quality of ZMR wafers. It was thus appropriate to extend Lincoln's demonstrated WS capability to SOI. While ZMR material was more readily available to us, being produced in-house, one goal of the program was to compare ZMR material with SIMOX (Separation by IMplanted OXYgen) wafers.

6.2 TECHNICAL ISSUES

Previous IC fabrication with ZMR material was done in a separate laboratory which does not have WS capability, so a new process had to be developed in the WS facility. Several technical issues have been addressed in the development: SOI substrate defects, device isolation, gate oxide quality, conduction by parasitic channels, and the integration of hard oxide process techniques with a WS integration process. Substrate defects proved to be a critical aspect of the development, with the device results contributing to the considerable progress in ZMR quality which occurred over the duration of this project.

The technical approach taken was to adapt an existing 2- μ m CMOS process for SOI substrates and use an existing mask set as the test vehicle. Later, a redesign of an existing, proven WS circuit was used for further development. Processes were developed for both SIMOX and ZMR substrates to determine which SOI technology was optimum for WS integration, the relationship between SOI thickness and device performance, and the effects of isolation techniques on device leakage and radiation resistance. Most work used SiO₂ as the gate dielectric, with incorporation of nitrided oxide (discussed elsewhere in this report) to improve the total dose hardness just beginning.

6.3 EXPERIMENTAL RESULTS

6.3.1 Thickness

The desirable silicon layer thickness is a trade-off among a number of factors. If the film is too thick, then the source-drain dopants do not penetrate through to the buried oxide, which increases the diode leakage. This restriction was mitigated in our process by using a small amount of phosphorus along with the primary arsenic dopant. A thick film also may cause difficulty with step coverage. If the film is too thin, then the transistor channel and the parasitic back channel (adjacent to the buried oxide) are not isolated from one another, and total-dose radiation hardness is degraded. A thin film also requires more complex metallurgy in the contact regions.

These factors lead to an optimum thickness of about 300 nm. In the case of ZMR, as discussed elsewhere in this report, the silicon quality is better for films in the 500- to 1000-nm range, so a process was developed to reduce the thickness to the desired value by oxidation. The thinning procedure was found to accentuate nonuniformity and defects in the ZMR material, but films of good quality produced by this technique are now available commercially from Kopin Corporation. Such material is being used at present to build two of the FPP cells described elsewhere in this report.

SIMOX material is limited by the energy capability of high-dose oxygen ion implanters to a thickness of slightly over 200 nm. A thicker film can be grown epitaxially on top, but that is a considerable complication in processing, so all our experiments have used films of 200 nm or less. This material is also commercially available, from IBIS Corporation. We initially developed a process for 1400°C annealing to form the buried oxide and the high-quality silicon film, but since then the vendor has begun supplying annealed wafers of good quality.

In a separate project we are conducting research in the fabrication and processing of very thin SOI transistors, sufficiently thin that the body of the device is fully depleted. Transistors with high mobility, exceeding $800 \text{ cm}^2/\text{V-s}$, have been fabricated in films as thin as 60 nm. A technique for local thinning of the silicon film has been developed to ease the difficulty of making contact to very thin silicon and to minimize the sheet resistance of the source/drain regions.

6.3.2 Isolation

Individual transistors built in the silicon film must be isolated from one another by one of two methods, mesa etching or local oxidation. The latter is a simpler process, nearly the same as that used for conventional bulk silicon, so it was used in some early experiments. It was understood that the resulting thick SiO_2 on the transistor edges would degrade radiation hardness, so this was a temporary expedient to accelerate development. However, it was found that boron depletion during oxide growth reduced the doping at the lower corner of the device island to such an extent that an intolerable leakage path existed even without radiation. Therefore this line of research was abandoned, and all later experiments used mesa isolation.

In the mesa technique, islands of silicon are etched by a plasma procedure which produces slightly sloping sidewalls and exposes the buried oxide in the regions between device islands. The gate oxide is then grown to completely cover the device. The buried oxide is etched slightly – tens of nanometers – underneath the edges of the island during cleaning. This exposed corner of silicon, as well as the corresponding corner at the top of the island, were items of concern with respect to gate oxide quality. Coverage by the polysilicon gate over the steep, 300-nm-high step was also a concern. Techniques were developed which allowed good devices to be made such that, for example, one group of 7 wafers had a total of 19,008 test transistors with 100 percent yield of good gates.

6.3.3 Parasitic Channels and Threshold Control

In addition to the desired front channel, an SOI transistor has parasitic channels at the back surface and on the mesa edges. The threshold voltages and radiation characteristics of these channels are different from those of the front, so the low-risk approach is to make their thresholds high enough that they remain off during normal operation. The sidewalls of n-channel transistors tend to have lower threshold than the front, probably due to higher fixed positive charge. The low sidewall threshold appears electrically as a step in the subthreshold current (curve 1 of Figure 6-1), a device with a difference of about 1.3 V between the two thresholds.

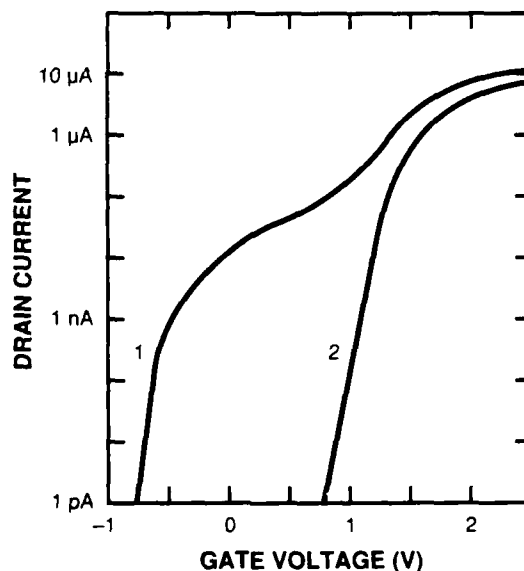


Figure 6-1. Drain current versus gate voltage for two $100\text{-} \times 100\text{-}\mu\text{m}^2$ n-channel ZMR transistors, one without sidewall doping (curve 1), and the other with (curve 2).

A process was developed to raise the n-channel sidewall threshold by increasing the boron doping of the island edges. Before etching the islands, boron is implanted into the n-channel field

areas and then diffused long enough to penetrate laterally beyond the edge of the etching mask. Curve 2 of Figure 6-1 illustrates such a sidewall-doped device. An alternative, simpler process is to dope after etching by ion implantation at an angle in the vicinity of 45° , but that capability is not presently available in our laboratory. Radiation effects are discussed in Section 6.3.4.

A parasitic back channel also exists for which the silicon wafer acts as the gate. To minimize radiation effects in the thick buried oxide, it is desirable to apply a negative voltage to the wafer relative to the circuitry. The wafer must be negative relative to the sources of the n-channel transistors, which typically are at 0 V. Sources of p-channel transistors typically are at 5 V, so the wafer voltage is at least 5 V in the direction of turning on the p-channel devices, and the back-channel threshold of those devices must be rather high. This requires relatively high n-type doping near the back interface. High doping, however, tends to make the front-channel threshold too high (too negative), so a substantial degree of counterdoping near the front interface is required. Such counterdoping is conventionally done by ion implantation through the gate oxide followed by minimal thermal processing, in order to avoid diffusion of the dopant. Unfortunately, such an implant severely degrades the radiation hardness of the gate dielectric. For good hardness the threshold-adjust implant must be done before growth of the gate oxide, and substantial boron diffusion occurs during the oxidation. Figure 6-2 is a SUPREM simulation of one such process which illustrates that the capability of threshold control by dopant profiling is limited in these devices. As a result the p-channel devices which we have built to date are fully depleted in normal operation, a condition which reduces the radiation hardness. We have recently developed a boron-doped polysilicon gate process which, by shifting the threshold without requiring counterdoping, is expected to allow *non-depleted* operation.

6.3.4 Radiation Hardness

Back channel hardness of 10 Mrad(Si) was demonstrated for ZMR material prior to this program[12]. Our recent results have confirmed that with negative voltage on the wafer, the back channel threshold remains high.

A rad-hard gate oxide process tolerant to 1 Mrad(Si), based on the Sandia Mod-B process[13], was developed using bulk silicon devices. Hardness of the edge channels on ZMR devices, however, was found to be relatively poor. (An early experiment indicated hard sidewalls, but later tests showed that result to have been in error.) We are presently building devices to test a sidewall process developed at Sandia which has been shown[14] to produce very high levels of hardening.

Nitrided oxide devices built in SOI material show hardness the same as the very high levels demonstrated in bulk devices, but they also suffer from the relatively soft sidewall. The Sandia sidewall hardening process is equally applicable to this gate dielectric. Producing the correct threshold voltages with nitrided oxide will require the use of boron-doped polysilicon gates.

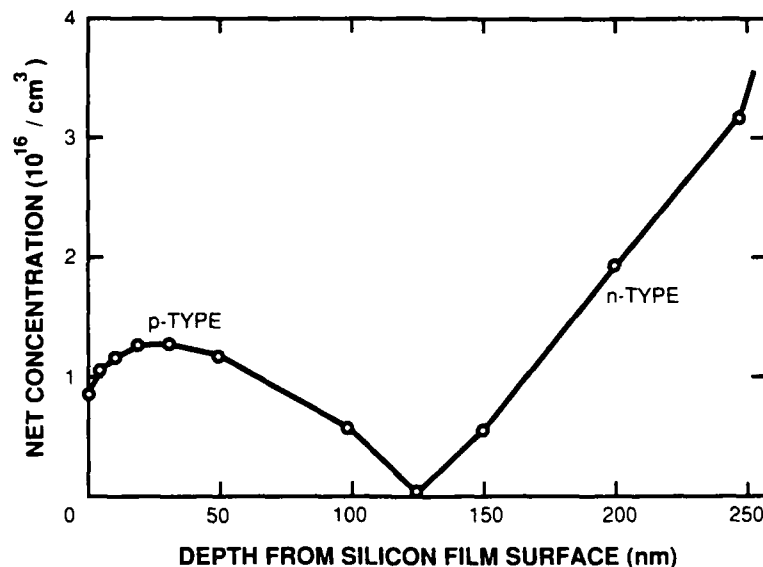


Figure 6-2. Simulated doping profile of a p-channel transistor body with deep arsenic and shallow boron implants.

6.3.5 Material Quality

The quality of SOI material was found to differ from bulk silicon in two ways, which are significant for this project. First, the gate oxide breakdown voltage and charge-to-breakdown are not as good, due either to protrusions or to impurities in the silicon. As discussed by Lee and Burns [15], oxides grown on older ZMR material, which contained a high density of low-angle grain boundaries (subboundaries), was substantially inferior. For example, the voltage for conduction of $1 \mu\text{A}/\text{cm}^2$ through a 37-nm oxide was 17 V for the early ZMR, 14 V for SIMOX, and 27 V for oxide grown on bulk silicon. ZMR films made by the presently recognized process, recrystallization of a relatively thick film followed by oxidation thinning, are much better. For example, Lee and Chen [16] found the breakdown voltage of a 37-nm oxide grown on such a subboundary-free film to be 80 to 90 percent that for an oxide on bulk silicon.

Second, while subboundaries have been shown to have little effect on transistor mobility[17], they appear to cause leakage paths between source and drain. An example is shown in Figure 6-3. The plateau of current at about $1 \mu\text{A}$ is only weakly dependent on both gate voltage (shown) and wafer voltage, which means that it is not caused by any of the parasitic channels. The drain diode reverse current measured on devices with a body contact is much lower than the plateau, so the leakage is not due to the diodes. This current appears rarely, if ever, in p-channel devices, but occurs in the order of 1 percent of n-channel transistors built in the older ZMR material. It has not been observed in devices built in subboundary-free ZMR or in SIMOX. While it has not been possible to demonstrate a one-to-one correlation between subboundaries and this leakage, it is

believed to result from either fast diffusion of source-drain dopant along the defect or an inversion path created by charges in the subboundary.

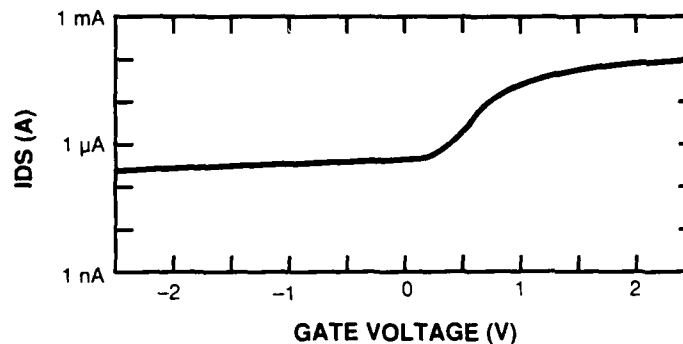


Figure 6-3. Subthreshold current plot of a $45 \times 3\text{-}\mu\text{m}^2$ transistor illustrating the leakage current believed to be associated with subboundaries in older ZMR material.

6.3.6 Integrated Circuit Results

The first SOI integrated circuits built in this program were a multiply accumulator and a parallel-serial converter comprising about 1000 and 5000 transistors, respectively. They are the cells of a WS IC, the multiply-accumulate array (MAA). They are CMOS circuits utilizing two-level metal with $3\text{-}\mu\text{m}$ minimum dimensions. Most importantly, they use dynamic logic, in which information is retained as charge stored on isolated nodes. For such circuits, transistor leakage must be less than $1\text{ }\mu\text{A}$ in order to store the charge for sufficient time. The leakage mechanism described above caused the yield of these circuits to be very low, but fully functional circuits of both types were built which operated at about 10 MHz.

Wafer scale MAA circuits were built, and all the issues important for laser restructuring were addressed although the cell yield was too low to actually program a wafer. No difficulty with wafer-length wiring was encountered. Vertical-link connections were successfully made with the laser, and it was shown that the laser linking process was the same as on bulk silicon wafers. Optical probing, a test technique required for restructuring, is different in SOI than in bulk, but it was shown that a transistor with its gate connected to its drain functions properly in place of the diode which is used in bulk circuits.

The circuits designed for the FPP use static logic, as is appropriate for an environment where radiation-induced currents are expected. Thus, not only will the incidence of high-leakage transistors be low because of improved material quality, but the effect of an occasional substandard transistor will be small. Two of the five FPP cells are in fabrication now for the continued development of SOI processing.

6.4 SUMMARY AND OUTSTANDING ISSUES

An SOI integrated circuit fabrication process suitable for either ZMR or SIMOX material was developed, and all issues peculiar to WS circuits on SOI were addressed. The circuit chosen for the initial WS experiment was particularly sensitive to transistor leakage current, and the early ZMR material which was used suffered from such leakage, so the yield was poor. Presently available material does not exhibit high leakage, and the FPP circuits are relatively insensitive to it, so much better yield is expected for the circuits now in fabrication.

Material deficiencies identified in this project appear, based on test device results, to have been solved, so that the probability of success in building large integrated circuits in either ZMR or SIMOX material is acceptable. Gate oxide quality and channel leakage are good, and mesa sidewalls can be doped to raise the parasitic threshold. Both types of SOI material are commercially available.

In continuing work, funded from other sources, testing the sidewall hardening process reported by Sandia will be completed, and it will be applied to nitrided oxide gates. Boron-doped polysilicon will be incorporated as the gate electrode so that correct threshold voltages can be produced while maintaining maximum hardness and preventing full depletion of the p-channel transistors. The two FPP cells currently in fabrication will be completed, and these designs will continue to be used to optimize the process and demonstrate yield, including the comparison of ZMR and SIMOX.

7. RADIATION-HARDENED REOXIDIZED NITRIDED OXIDES

The use of SOI eliminates or greatly suppresses most of the circuit failure modes associated with radiation, but it has no direct effect on total dose damage to the gate dielectric. A program is in progress at Lincoln Laboratory to develop an extremely hard gate dielectric by the nitridation of silicon dioxide. While separately funded, this program has been closely allied with the Radiation-Hardened Wafer Scale Program, and its results will be incorporated in future SOI circuits. We have demonstrated a 37-nm dielectric which exhibits zero interface state increase and only -1.35 -V threshold voltage shift after 100 Mrad(Si), very high resistance to channel hot carrier stress, and a factor-of-seven improvement in charge-to-breakdown (Q_{bd}) over conventional oxide.

This dielectric is produced by first growing a conventional oxide of the desired thickness and then, in the same furnace tube, partially converting it to a nitride by exposure to ammonia (nitridation), followed by a second oxidation. The process can easily be incorporated in a typical fabrication sequence.

Midgap interface state density versus dose in capacitors irradiated with positive 5-V bias applied to the gate is plotted in Figure 7-1. The suppression of interface state generation in nitrided and reoxidized nitrided oxides is clearly illustrated. In capacitors with nitrided oxide as the gate dielectric, midgap D_{it} increased by less than $1 \times 10^{11}/\text{cm}^2$ eV after 100 Mrad(Si). In reoxidized nitrided oxide capacitors, no interface state generation was measured (within the resolution of the measurement technique: $< 1 \times 10^{10}/\text{cm}^2$ eV). We have also found that no interface states are generated in these dielectrics when they are subjected to high-field current stress [18]. Finally, we note that the substantial numbers of interface states generated in the oxides become even larger with time after irradiation if the positive gate bias is maintained. Bias annealing did not produce any change in D_{it} in the reoxidized nitrided oxide.

Midgap voltage shift versus applied field during irradiation is plotted in Figure 7-2 for nitrided oxide and a nonoptimum reoxidized nitrided oxide after 5 Mrad(Si), and for conventional oxide after 1 Mrad(Si). Oxide capacitors exhibit the well-known behavior of greater midgap voltage shift with positive applied gate bias. This is because positive bias sweeps radiation-generated holes toward the Si/SiO₂ interface, where the hole traps are known to be concentrated and where trapped holes will have the greatest moment in shifting V_{mg} . In the reoxidized nitrided oxide, however, the opposite behavior is observed: voltage shift is greater when negative bias is applied to the gate during irradiation. An etch-off experiment [19] demonstrated that this unusual behavior is due to a relatively high concentration of hole traps very close to the gate. The curve labeled "3.6-nm etch-off" in Figure 7-2 represents data from capacitors fabricated with the same reoxidized nitrided oxide but after a dilute hydrofluoric acid etch had removed 3.6 nm of the gate dielectric. A dramatic reduction in negative bias voltage shift is observed.

We have found that this large negative bias shift can be eliminated by optimization of the device fabrication process. Midgap voltage shifts for capacitors fabricated with this optimum process are plotted in Figure 7-3 for total doses up to 100 Mrad(Si).

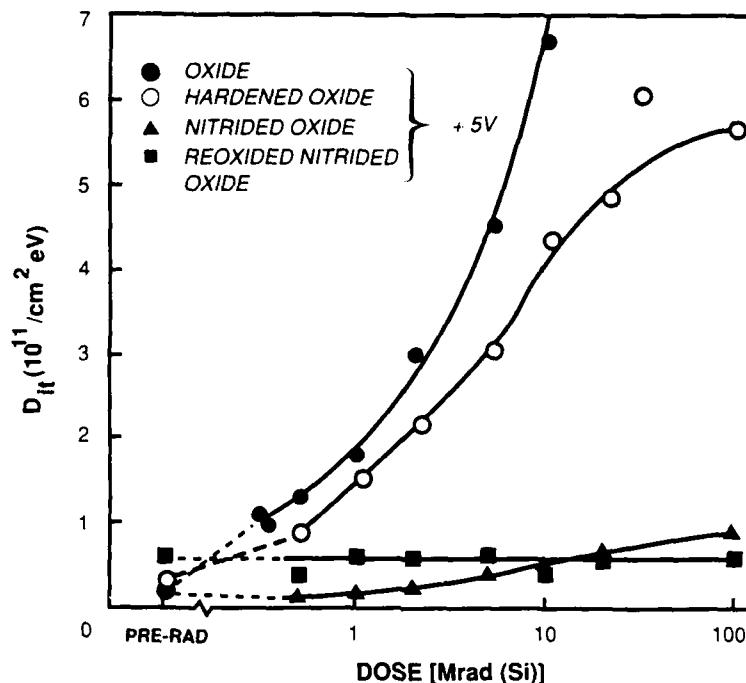


Figure 7-1. Midgap interface state density versus 10-keV X-ray dose in capacitors with hard and soft oxide, nitrated oxide, and reoxidized nitrated oxide as the gate dielectric. +5 V was applied to the gate during irradiation.

Voltage shifts for the nitrated oxide without reoxidation and for the hard oxide are also shown. ΔV_{mg} for the reoxidized nitrated oxide under either polarity of gate bias is less than -1.35 V after 100 Mrad(Si), substantially less than that of the hard oxide devices. It is interesting to note that the behavior of the nitrated oxide without reoxidation is very different. Although midgap voltage shift is very small at low doses, it continues to increase with dose, surpassing that of the reoxidized devices at <10 Mrad, and approaching that of the hard oxide devices at 100 Mrad. We speculate that this behavior is due to a relatively high concentration of small capture cross-section traps in the nitrated oxide.

Although we have not yet fabricated transistors with the optimized process, radiation testing of transistors with a suboptimum process has demonstrated that hardness can be preserved through the full CMOS fabrication process. Midgap voltage shifts for p-channel transistors with channel length and width of 2.5 and 10 μm , respectively, are plotted in Figure 7-4.

Transistors with hard oxide as the gate dielectric were also tested for comparison. Although the suboptimum processing resulted in a relatively large negative bias shift in midgap voltage for the reoxidized nitrated oxide devices, ΔV_{mg} for either bias is still less than in hard oxide devices for total doses greater than 1 Mrad(Si). In addition, no change in subthreshold slope was measured in

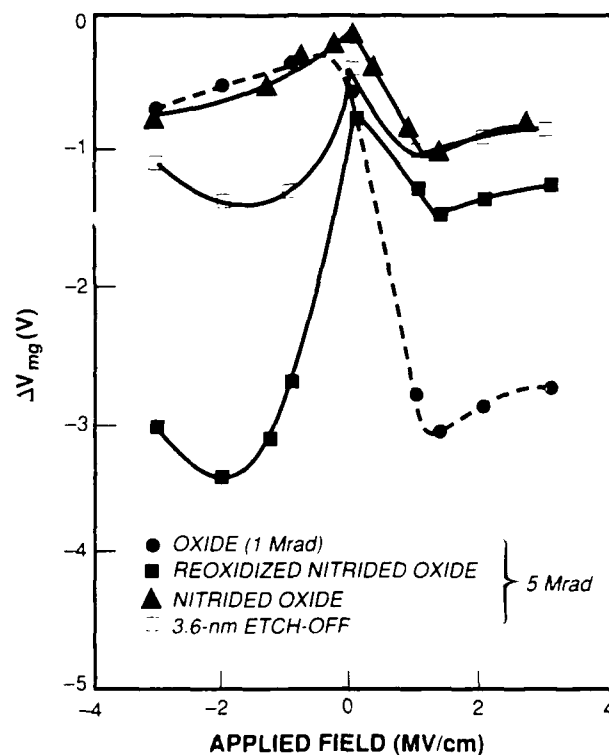


Figure 7-2. Midgap voltage shift versus applied oxide field in soft oxide capacitors irradiated to 1 Mrad(Si) and nitrided oxide and reoxidized nitrided oxide capacitors irradiated to 5 Mrad. The curve labeled "3.6 nm etch-off" refers to a reoxidized nitrided oxide which was etched in hydrofluoric acid to remove 3.6 nm of material prior to gate deposition.

reoxidized nitrided oxide devices, reflecting the complete suppression of interface state generation. Inversion layer mobilities normalized to the prerad oxide value are plotted in Figure 7-5.

Although μ_{eff} is initially lower in the reoxidized nitrided oxide devices [20], no significant degradation is observed even after 100 Mrad(Si). Mobility in the hard oxide devices degrades considerably, such that after 5 Mrad(Si) μ_{eff} is lower than in the reoxidized nitrided oxide devices.

Dielectric durability was studied by charge-to-breakdown measurements, in which capacitors were stressed in a constant current mode at 0.01 A/cm^2 with the gate positive. Although Q_{bd} for the nitrided oxide without reoxidation was smaller by a factor of 4 compared to oxide (7 versus 30 C/cm^2), Q_{bd} for reoxidized nitrided oxide was 200 C/cm^2 , greater by almost a factor of 7.

Transconductance degradation due to channel hot carrier stress is an important problem in

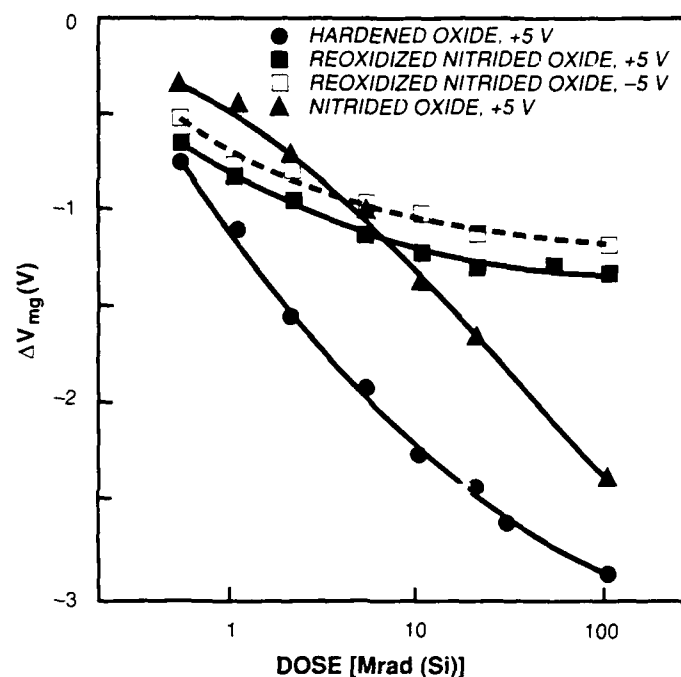


Figure 7-3. Midgap voltage shift versus X-ray dose for capacitors with hard oxide, nitrided oxide, and an optimized reoxidized nitrided oxide as the gate dielectric.

short channel devices. We investigated the hot carrier resistance of reoxidized nitrided oxides. N-channel transistors with channel length and width of 1.3 and 10 μm , respectively, were stressed at drain voltages between 6.5 and 8.5 V and gate voltage corresponding to peak substrate current, which occurs at $V_g \approx V_{d/2}$, and is known to be the worst-case stress condition. Peak linear region transconductance was measured in reverse mode with $V_s = 50\text{mV}$ before stress and after 5000 s. Percent degradation versus peak substrate current is plotted in Figure 7-6.

At similar values of I_{sub} , degradation in reoxidized nitrided oxide devices is lower by a factor of ten than that in oxide devices. Shifts in the extrapolated threshold voltage were negligible for both dielectrics. The superior performance of reoxidized nitrided oxide is believed to be due to the suppression of interface state generation, as this is considered to be the dominant degradation mechanism of hot carrier stress.

In summary, our research program in reoxidized nitrided oxides has led to the development of a dielectric which exhibits greatly improved resistance to ionizing radiation and channel hot carrier stress. We believe that reoxidized nitrided oxide shows great potential for use in small geometry CMOS circuits for applications in radiation environments.

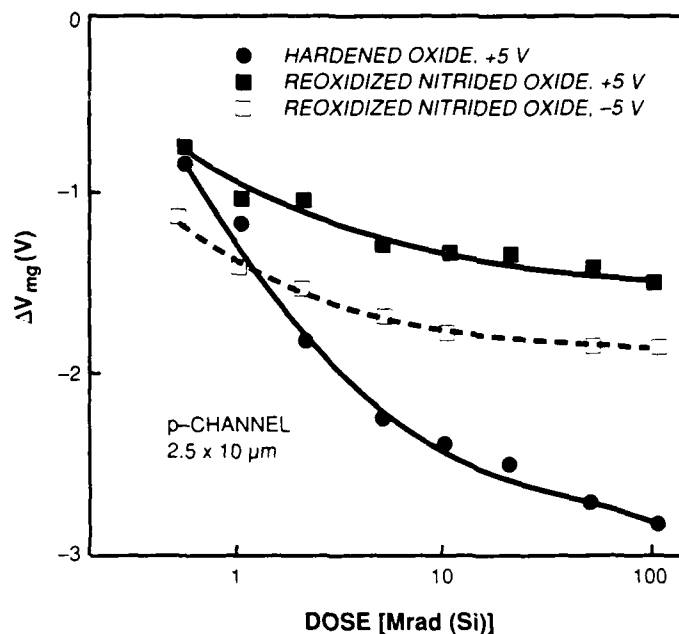


Figure 7-4. Midgap voltage shift versus X-ray dose in p-channel transistors with hard oxide and a suboptimum reoxidized nitrided oxide. ± 5 V was applied to the gate during irradiation. Source, drain and substrate were grounded.

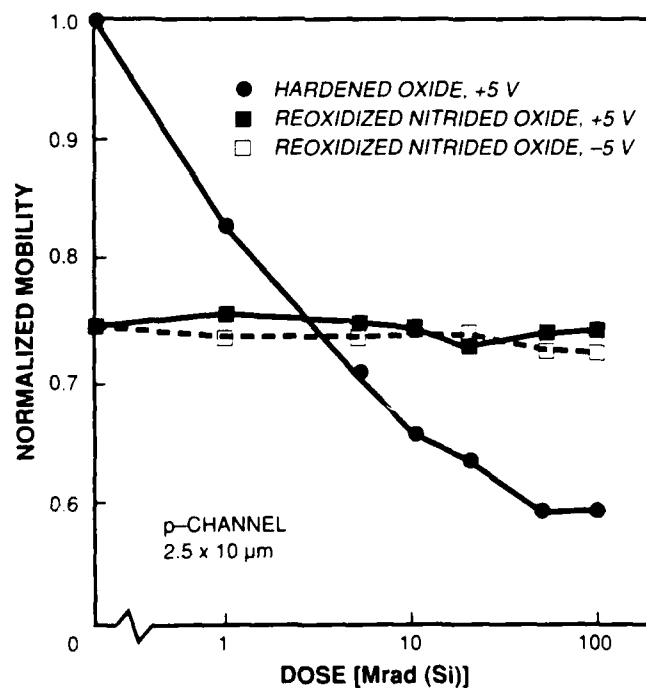


Figure 7-5. Normalized inversion layer mobilities of p-channel transistors as described in Figure 7-4.

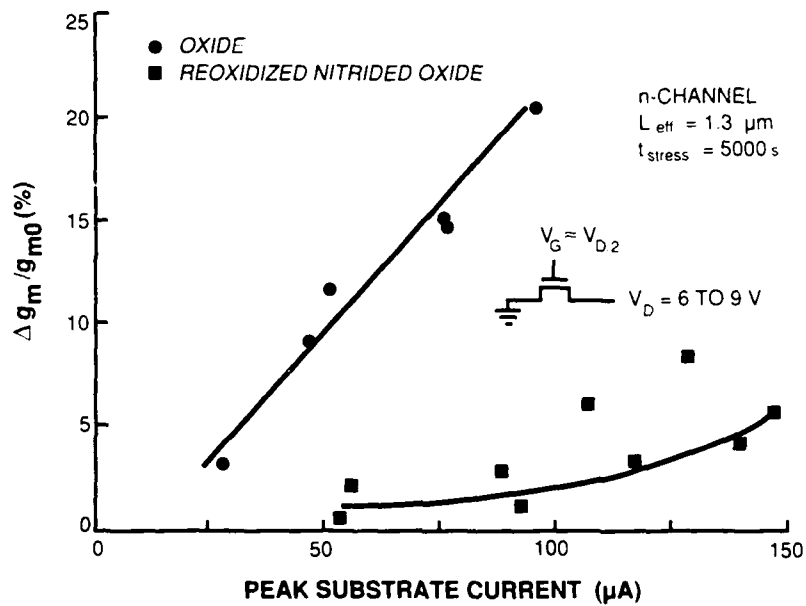


Figure 7-6. Percent peak linear region transconductance degradation versus peak substrate current after 5000-s channel hot carrier stressing of n-channel transistors with soft oxide and reoxidized nitrided oxide.

127224-29

8. CONCLUSIONS

FPP design is complete. The five circuits have been fabricated in a bulk CMOS process with very high yields and operate at higher speed than the design goal of 16 MHz. The WS layout has not been finished but completion would be very straightforward. On a 3-in substrate with a 3- μ m process, enough cells would be fabricated in a 41 \times 45 mm rectangle to build five processors with 2X redundancy for the larger cells and 1.6X for the smaller *delay* cell. On bulk silicon there is very high confidence that the system could be fabricated with high yield and operate as designed. On a 5-in bulk silicon wafer with 2 μ m processing, enough cells could be built to place 40 or 50 processors on a wafer.

A fabrication process for CMOS circuits on SOI wafers from either the ZMR or SIMOX process was developed. Two of the FPP cells will be fabricated on SOI wafers, and there appear to be no problems which would hinder fabrication of WS circuits.

A new system for fabricating SOI films by ZMR was designed, fabricated, and tested. Promising results were obtained in producing films on larger substrates than previously possible. Material can now be produced which is free of subgrain boundaries.

In a related program, a reoxidized nitrided oxide film has been developed which exhibits greatly improved resistance to channel hot carrier stress and radiation hardness of 100 Mrad(Si). This gate dielectric will be incorporated into the SOI fabrication process.

REFERENCES

1. J. Raffel, A.H. Anderson, G.H. Chapman, "Laser Restructurable Technology and Design," Ch. 7 in *Wafer Scale Integration*, Earl Swartzlander, ed., pp. 319-363, Kluwer Academic Publishers, Boston, MA (1989).
2. P.C. Trepagnier, "Design of a Wafer-Scale Focal Plane Processor" Technical Report 814, MIT Lincoln Laboratory (15 September 1988).
3. A.H. Anderson, R. Berger, K.H. Konkle, F.M. Rhodes, "RVLSI Applications and Physical Design," *Proc. of International Conference on Wafer Scale Integration*, San Francisco, CA. IEEE Computer Society Press (January 1989).
4. A good introduction to SOI fabrication is given in a special issue of *IEEE Circuits Devices Mag.* Vol. 3 (July 1987).
5. J.K. Ousterhout, G.T. Hamachi, R.N. Mayo, W.S. Scott, and G.S. Taylor, "Magic: A VLSI Layout System," *Proc 21st Design Automation Conference*, pp. 152-159 (1984).
6. R.E. Bryant, D. Beatty, K. Cho, and T. Sheffler, "COSMOS: A Compiled Simulator for MOS Circuits," *Proc. 24th Design Automation Conference*, pp. 9-16 (1987).
7. R. Frankel, J.J. Hunt, M. Van Alstyne, G. Young, "SLASH-An RVLSI CAD System," *Proc. of International Conference on Wafer Scale Integration*, San Francisco, CA. IEEE Computer Society Press (January 1989).
8. S.L. Garverick and E.A. Pierce, "A Single Wafer 16-Point 16-MHz FFT Processor," *Proc. of the Custom Integrated Circuits Conference*, 244-248 (1983).
9. R.F. Lyon, "Two's Complement Pipeline Multipliers," *IEEE Trans. Commun.*, COM-24, 418-25 (1976).
10. N. Ohwada, T. Kimura, and M. Doken, "LSI's for Digital Signal Processing," *IEEE J. Solid-State Circuits*, SC-14, 214-20 (1979).
11. C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, New York, 78-79 (1980).
12. B-Y. Tsaur, R.W. Mountain, C.K. Chen, G.W. Turner, and J.C.C. Fan, "Effects of Ionizing Radiation on SOI/CMOS Devices Fabricated in Zone-Melting-Recrystallized Si Films on SiO₂" *IEEE Electron Device Lett.*, EDL-5, 238 (1984).
13. P.S. Winokur, E.B. Errett, D.M. Fleetwood, P.V. Dressendorfer, and D.C. Turpin, "Optimizing and Controlling the Radiation Hardness of a Silicon Gate CMOS Process," *IEEE Trans. Nucl. Sci.*, NS-32, 3954 (1985).

14. S.S. Tsao, D.M. Fleetwood, and H.T. Weaver, "Radiation-Tolerant, Sidewall-Hardened SOI/MOS Transistors," *IEEE Trans. Nucl. Sci.*, NS-34, 1686-1691 (1987).
15. C-T. Lee and J.A. Burns, "Study of Gate Oxide Leakage and Charge Trapping in ZMR and SIMOX SOI MOSFETs," *IEEE Electron Device Lett.*, EDL-9, 235-237 (1988).
16. C-T. Lee and C.K. Chen, "Defect-related dielectric breakdown, charge trapping, and interface-state generation of gate oxides grown on zone-melting-recrystallized silicon-on-insulator films," *J. Appl. Phys.* Vol. 65, 646-650 (1989).
17. B-Y. Tsaur, J.C.C. Fan, M.W. Geis, D.J. Silversmith, and R.W. Mountain, "Effects of Subgrain Boundaries on Carrier Transport in Zone-Melting-Recrystallized Si Films on SiO₂-Coated Si Substrates," *IEEE Electron Device Lett.*, EDL-3, 79-82 (1982).
18. G. Dunn, "Generation of Interface States in Nitrided Oxide Gate Dielectrics by Ionizing Radiation and Fowler-Nordheim Stress," *Appl. Phys. Lett.* Vol. 53, 1650 (1988).
19. G. Dunn, "Hole Trapping in Reoxidized Nitrided Silicon Dioxide," *Appl. Phys.*, in press, 1989.
20. M.A. Schmidt, F.L. Terry, B.P. Mathur and S.D. Senturia, "Inversion Layer Mobility of MOSFET's with Nitrided Oxide Gate Dielectrics," *IEEE Trans. Electron Devices*, Vol. 35, 1627 (1988).

APPENDIX
Design of a Wafer-Scale Focal Plane Processor

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

DESIGN OF A WAFER-SCALE FOCAL PLANE PROCESSOR

P.C. TREPAGNIER
Group 27

TECHNICAL REPORT 814

15 SEPTEMBER 1988

Approved for public release; distribution unlimited.

LEXINGTON

MASSACHUSETTS

ABSTRACT

This report describes a wafer-scale design for an infrared focal plane processor (FPP) to operate in a space environment. The functions of a generic focal plane processor are described, followed by a detailed discussion of a design to be implemented in RVLSI wafer-scale technology for a space-based application. A prototype of this processor (PFPP) will actually be fabricated in rad-hard silicon-on-insulator 3- μm technology. Finally, the question of reliability is explored, and a philosophy of fault-tolerance is presented which will lead to a reasonable probability of success over a five-year lifetime.

TABLE OF CONTENTS

ABSTRACT	iii
LIST OF ILLUSTRATIONS	vi
LIST OF TABLES	vii
ACKNOWLEDGEMENTS	ix
1. INTRODUCTION	1
1.1 Scanning Arrays	1
1.2 Focal Plane Processor	1
1.3 Fault Tolerance	2
1.4 Wafer Scale Integration	3
1.5 Organization of Report	3
2. DESIGN OF A PROTOTYPE FOCAL PLANE PROCESSOR	5
2.1 Wafer-Level Description	5
2.2 Fault-Tolerance Input	7
2.3 Piecewise Linear Calibration	7
2.4 Time Alignment	10
2.5 Gamma Circumvention and TDI Summation	11
2.6 Matched Filter and Detector	14
3. IMPLEMENTATION ISSUES	19
3.1 Area Estimates	19
3.2 Serial versus Parallel Arithmetic	19
3.3 Predicted versus Actual Area	23

4. RELIABILITY ESTIMATES	25
4.1 MIL-HDBK-217E	25
4.2 Processor Element Reliability Estimate	26
4.3 FPP Reliability Estimate	26
4.4 Influence of Pairwise Interactions	27
5. CONCLUSION	31
APPENDIX A - MAXIMUM LIKELIHOOD APPROACH TO GAMMA CIRCUMVEN- TION	33
A.1 Maximum Likelihood Model	33
A.2 Illustrative Example and Comparison with SATDI	34
APPENDIX B - ROUND OFF IN TDI SUMMATION	39
REFERENCES	42

LIST OF ILLUSTRATIONS

Figure No.		Page
2-1	Downsized scanning array	6
2-2	Schematic wafer layout	8
2-3	Input and calibration cell	9
2-4	Time delay cell	11
2-5	Gamma circumvention and TDI summation cell	13
2-6	Matched filter and detection cell	16
2-7	FIR filters	17
3-1	Schematic area allocation on PFPP wafer – bit parallel	21
3-2	Schematic area allocation on PFPP wafer – bit serial	22
4-1	Processor reliability	28
4-2	Processor reliability with nearest-neighbor interaction	29
4-3	Effects of PE reliability	30
A-1	Probability density function	35
A-2	Log likelihood ratio	35
A-3	Receiver operating characteristic	37
B-1	Detail of proposed TDI summation circuit	39
B-2	Low background, (a) high gamma and (b) low gamma simulation.	40
B-3	High background, high gamma simulation.	41

LIST OF TABLES

Table No.		Page
3-1	Component area estimates	19
3-2	Area allocation for a PE - bit parallel	21
3-3	Area allocation for a PE - bit serial	22
3-4	Estimated versus actual cell area	23

ACKNOWLEDGEMENTS

The author, who was completely ignorant of wafer-scale technology when this work began, benefited greatly from discussions with the members of Group 23, who are doing the design and fabrication of the wafer described in this work as well as previous monolithic restructurable VLSI wafers. The contributions of A. H. Anderson, C. E. Woodward, and K. H. Konkle deserve particular mention.

A great debt is owed to Steve Pohlig of Group 27, who did much of the early conceptual design work on large-scale FPPs before the present author came to Lincoln.

Finally, it is a pleasure to acknowledge the many conversations with A. E. Filip, who shepherded the project from its inception.

DESIGN OF A WAFER-SCALE FOCAL PLANE PROCESSOR

1. INTRODUCTION

1.1 SCANNING ARRAYS

Consider a generic scanning infrared sensor, consisting of a detector array with n rows and k time delay integration (TDI) columns. (The entire arrangement may then be duplicated for each of m color bands. These will be ignored hereafter for the sake of simplicity.) One can imagine this array scanning horizontally across an image in order to form a two-dimensional picture. It moves horizontally by one column every dwell, and in addition is oversampled, typically by a factor of three, so that the entire set of detectors is read out three times per dwell. Data from a column with TDI position k must be delayed $k - 1$ of these dwells before being added to subsequent data from the same row in order to perform the time alignment needed for integration.

1.2 FOCAL PLANE PROCESSOR

The focal plane processor (FPP), also known as a time dependent processor, is responsible for the initial signal processing of data from an array of photodetectors. From a computational point of view, the initial focal plane processing is characterized by two salient points: (a) the input data stream is massively parallel: each detector in the scanning array is sampled after every dwell time and is treated essentially identically, and (b) the algorithms applied to each detector sample are relatively simple and well-understood. These two points taken together favor a hardwired, single instruction multiple data (SIMD) architecture for the FPP. This architecture, together with the requirements of low power consumption, low weight, and high reliability imposed by a space environment, makes wafer scale integration (WSI) a natural choice for the processor technology. Nonetheless, even the relatively simple processing requirements of the FPP impose a higher degree of internal differentiation on the WSI processor (i.e., more cell types) than has previously been demonstrated. Design of such a WSI processor is a nontrivial task, and represents the subject of this report.

The functions of the FPP may now be discussed in greater detail. The incoming data must be calibrated to correct for responsivity differences among detectors, and samples which have been corrupted by the effects of γ radiation need to be recognized and discarded. Following that, two other signal processing functions, time-delay integration and matched filtering and threshold detection, must be performed. At this point, the object dependent processor (ODP), whose load depends on the number of objects over threshold, takes over. These four major functional units are described briefly in the order in which the data pass through them.

1.2.1 Calibration

Each pixel in the detector array will have a slightly different dark current and responsivity, which must be corrected. If this function has not been implemented in the analog front end, it is

handled in the FPP via an addition and multiplication. In principle, nonlinear responsivities could also be calibrated out. This is rarely done in practice due to the difficulty of finding appropriate calibration standards.

1.2.2 Time Alignment

The earlier columns of the scanning array must be delayed before being added to later columns. This function, which would be performed by a CCD shift register in analog implementations, is implemented digitally as a circular buffer.

1.2.3 Gamma Circumvention

The detection of γ -affected data is very much like a CFAR detector, where the threshold is set to a certain number of standard deviations beyond the mean. A current estimate of the mean and standard deviation of the signal is obtained using various semiheuristic methods, and the ensemble of TDI samples corresponding to a given point is compared with a threshold based on this estimate. Samples above this threshold are assumed to be contaminated by γ -induced electrons and are discarded. The remaining samples are then averaged together to form the TDI output.

1.2.4 Matched Filter and Detector

The output of time alignment is then run through an FIR filter which compensates for the combined effect of oversampling and the point spread function of the optics. In the simplest implementation, the detector is simply a comparator. More sophisticated FPPs may incorporate more complicated circuitry, e.g., Laplacian filters to remove nuclear background effects.

1.3 FAULT TOLERANCE

The goal of a five-year mission lifetime, combined with the expected reliability of wafer-scale circuits, imposes a fault-tolerant structure on the design. The approach taken here is to have redundant circuit elements which may be switched in as needed via multiplexors. There is a design tradeoff to be made on the size of these fault-tolerant elements - too small, and the switching circuitry becomes cumbersome; too large, and the probability of and penalty for failure both become excessive.

As will be seen below, this tradeoff was one of the factors influencing the choice of lower-capability serial arithmetic processors, rather than higher-capability parallel ones. The fault tolerant unit was then chosen to be a complete processing element (PE), comprising all four functional units.

1.4 WAFER SCALE INTEGRATION

Design of an FPP to be realized in wafer scale technology must take into account the requirements of this technology. Chiefly, this means that it must be possible to lay out the processor on a wafer, and that the processor must be manufacturable with a reasonable yield.

1.4.1 Serial versus Parallel Arithmetic

The layout problem became evident early in the consideration of a parallel processor. Since the processor was designed for 12-bit arithmetic, utilizing a 35- μ m wire pitch resulted in each bus being 0.4 mm wide. The combination of a fault-tolerant architecture and the requirement for processing parallel TDI stages leads naturally to a design in which several buses lie side by side. The resulting "Los Angeles effect" produces a wafer in which buses are a significant fraction of the total area (see Section 3.2.) This fact led to the consideration of nibble-wide buses. One-bit nibbles were rapidly realized to be most appropriate, at least in the near term.

1.4.2 Defect Tolerance

Any process will have a small number of manufacturing defects. A circuit containing as many elements as a wafer-scale processor will have a yield approaching zero unless a way is found to correct the defects after manufacture. In the restructurable VLSI processes, redundant elements called restructurable cells are laid down. These are then connected together after testing[9]. Hence, any design must include identification of suitable restructurable cells. These cells must be relatively small ($< 15,000$ transistors) so that their yield is good, yet be common and few in type to simplify design and mask production. Ideally, they should bear some simple relationship to the functions of the processor. All these goals are furthered by an architecture based on a multitude of low-capability serial elements, rather than a few higher-capability parallel ones. In particular, we find that the restructurable cells can be just the four functional units discussed in Section 1.2.*

1.5 ORGANIZATION OF REPORT

This report is divided into five sections. The present section introduces an FPP and its functions to those unfamiliar with one, and to identifies the principal issues that drive the design. Section 2 begins by presenting a set of strawman requirements for a space-based IR sensor. These

* Late in the design of the wafer, the gamma circumvention circuit was in fact split into two smaller parts, one for TDI summation (left side of Figure 2-5) and one for gamma threshold generation (right side). The change was made for producibility reasons; the full gamma cell would otherwise have been 40 percent larger than the next largest cell in the PFPP. For the purposes of this report, however, the two cells will be considered as one.

requirements motivate the design of the major functional units of a prototype wafer-scale FPP, which are described in some detail in the remainder of the section.

Sections 3 and 4 give a closer look at some of the critical design methodology. Section 3 describes in more detail the area calculations which illuminated the principal problem in the initial design of the WSI prototype FPP: getting enough processors on the wafer to ensure a reasonable probability of success. Success in this sense must embrace both initial yield (defect tolerance) and reliability in use (fault tolerance). The solution to this problem is the use of bit-serial arithmetic. Section 4 describes the part-stress-analysis approach [4] used to estimate the reliability of the PFPP and its subunits. Section 4 also presents a bottom-up calculation and rationale for the reliability parameters chosen. A redundant (M-of-N) processing element architecture is employed to achieve acceptable mission life given the expected subunit reliability.

Following the report conclusion, two appendices present more in-depth treatments of roundoff errors in TDI summing, and an alternate approach to infrared detection in the presence of gamma radiation.

2. DESIGN OF A PROTOTYPE FOCAL PLANE PROCESSOR

2.1 WAFER-LEVEL DESCRIPTION

The parameters for this design were based on the conclusions of a number of classified studies reflecting the projected requirements of the Space Surveillance and Tracking System (SSTS). The strawman sensor point design calls for a sensor of 20,000 rows and 5 TDI columns in each of 4 color bands. The detector array moves horizontally by 1 column every $28 \mu\text{s}$ and is oversampled by a factor of 4 in time, so that the entire set of detectors is read out every $7 \mu\text{s}$. The dwell time used in the TDI process is $28 \mu\text{s}$.^{*} A wafer scale (or any other) processor is unrealizable for this data rate ($4 \cdot 10^5 \text{ detectors} \times 1.4 \cdot 10^5 \text{ Hz} = 5.6 \cdot 10^{10} \text{ samples/s}$) in current technology, although one will eventually be feasible using one micron or smaller geometry and large wafers.

Instead, a prototype FPP (PFPP) was designed around a downsized scanning infrared sensor, shown in Figure 2-1. This sensor consists of a monochrome detector array with only 64 rows. However, the number of TDI columns and readout rate was retained from the strawman sensor, so that the PFPP maintains the essential design parameters of the complete sensor, but with 1250 times fewer processing elements. These PEs could then be proliferated on 6-inch wafers with $1\text{-}\mu\text{m}$ geometry, but need not be redesigned to accommodate the full strawman sensor point design.

The following list is a summary of the PFPP design, based on the above sensor description and assuming 3-inch SOI wafers with $3\text{-}\mu\text{m}$ design rules.

- (1) The processing of the 64 detector rows will be performed with a system using 2 wafers, which will contain 5 processor elements (PEs) – 4 working and 1 spare.
- (2) Each processor element processes data from 8 consecutive rows of the detector array.
- (3) Input data are assumed to be 12 bits long. This wordlength permits a mean background that is two orders of magnitude greater than the target signal. Three-percent precision (5 bits) is then possible on a signal that is one percent (7 bits) of the mean [1].
- (4) Eight detector rows are processed in the $7 \mu\text{s}$ sampling time requiring $7 \mu\text{s}/8 = 875 \text{ ns}$ per detector. In order to preserve full 12-bit accuracy throughout, the

^{*} Since the time of these SSTS studies, the space surveillance community has moved toward less aggressive sensor designs emphasizing near-to-intermediate term producibility. Typical integration times have become an order of magnitude or more longer and the number of detector elements has decreased, although the number of TDI stages have gone up somewhat. The design for this prototype processor, however, was frozen before these changes became effective. The principal effect of implementing the changes would be to make the FPP much more memory intensive, by increasing the size of calibration memories and delay buffers while reducing the number of PEs.

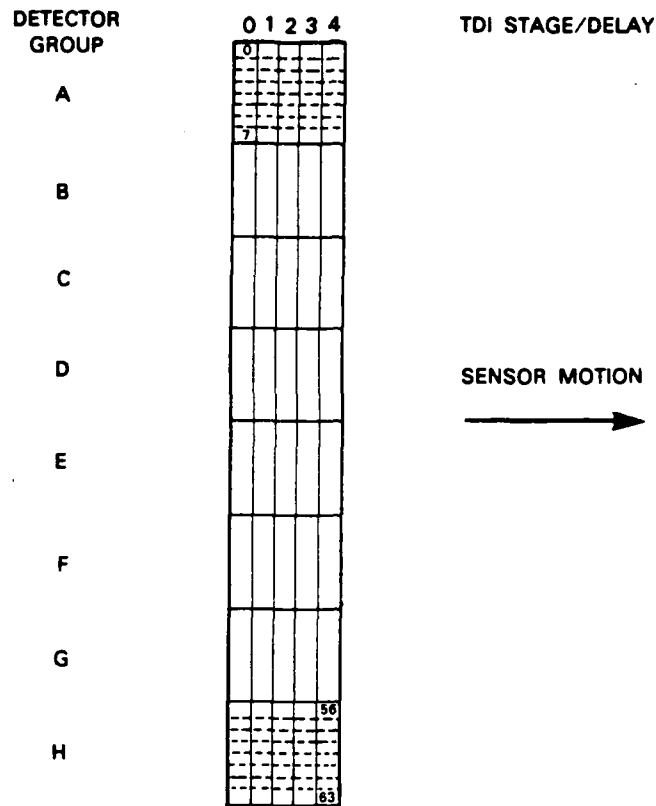


Figure 2-1. Downsized scanning array.

12-bit input data stream will be padded with 2 bits of leading zeros, providing for word growth in intermediate stages of processing. It will then be processed bit-serially. Thus, the processor clock will run at $875 \text{ ns}/14 = 62.57 \text{ ns/bit}$ (16 MHz).

- (5) Fault-tolerance is obtained by connecting the processor elements to the input and output buses through multiplexors, allowing any 2 of the 10 PEs to fail without loss of functionality.
- (6) Defect-tolerance is obtained by laying down a large number of PEs and piecing together good ones at restructuring time. Current area estimates indicate that 18 complete PEs could be laid down on a single wafer, however, fewer actually will be (see Section 3.3).

Note that only 5 of the possible 18 PEs per wafer are required to restructure the proposed system. Additional bussing and pinouts will be provided so that if yields are better than this initial conservative design goal requires, the wafer can be configured to handle a larger number of inputs.

The design still calls for a 2-wafer set in order to exercise the multiple wafer design concept which will eventually be required.

Figure 2-2 represents a quasi-geographical schematic layout of one wafer from a two-wafer processor set, with the lowest level of detail being four units – input mux/calibration/time alignment; gamma circumvention/TDI summation; matched filter/detector; and output mux. The fault-tolerant data busing is shown in detail on this figure, although the rest of the busing (e.g., off-wafer calibration, control logic, etc.) is not. This is to make the sparing strategy explicit, as well as show some of the complexity of the interconnect. Note that since the architecture is serial, all buses are only one bit wide.

2.2 FAULT-TOLERANCE INPUT

Figure 2-3 shows the input and calibration cell. Each input subunit is connected by a 4:1 multiplexor to any of 3 consecutive input signals (except for PEs on the ends of the chain) or a test pattern input. This arrangement permits any 2 PEs to fail at runtime, and to be replaced by their neighbors. Referring back to Figure 2-2, the 4 initially active PEs on the wafer are shown labeled A-D, corresponding to the array segments to which they are assigned. The second wafer (not shown), will have an identical set labeled E-H. The input subunits are also subscripted with the TDI stage to which they belong. Two spare PEs are provided, one at each end of the processor element chain, labeled X (shown in Figure 2-2) and Y (on the other wafer).

Figure 2-2 also shows output muxes for the PEs. This feature would make the sparing strategy transparent off-wafer; each output pin would always contain signals from the same input pixels. In the interest of simplicity, however, the output mux will not be implemented in the PFPP. Pins for every PE are present and the ODP will have to keep track of which are active.

In this design, the fault-tolerant atom is the whole PE. This approach, which simplifies the design concept, is made possible by the use of small low-capability serial processors. A parallel processor running at a similar clock rate, e.g., serving 96 rows instead of 8, would be too large to discard lightly.

2.3 PIECEWISE LINEAR CALIBRATION

Figure 2-3 shows the calibration circuit. The input data are processed by a piecewise linear approximation to a function which corrects for nonlinearity and nonuniformity in the detectors. There is a separate set of calibration coefficients for each of the 8 detectors assigned to a single PE. Each of these calibration functions has 4 linear segments. The appropriate slope and offset for the piecewise linear function are selected by addressing the coefficient memory with a combination of the 2 MSBs of the input data to indicate which of the 4 linear segments to use, and a counter to indicate which detector is being corrected.

The slope coefficients are stored with 10-bit accuracy. This length is sufficient to maintain the input accuracy, since each coefficient is applied over $\frac{1}{4}$ full range. Since the offset is 12 bits, each

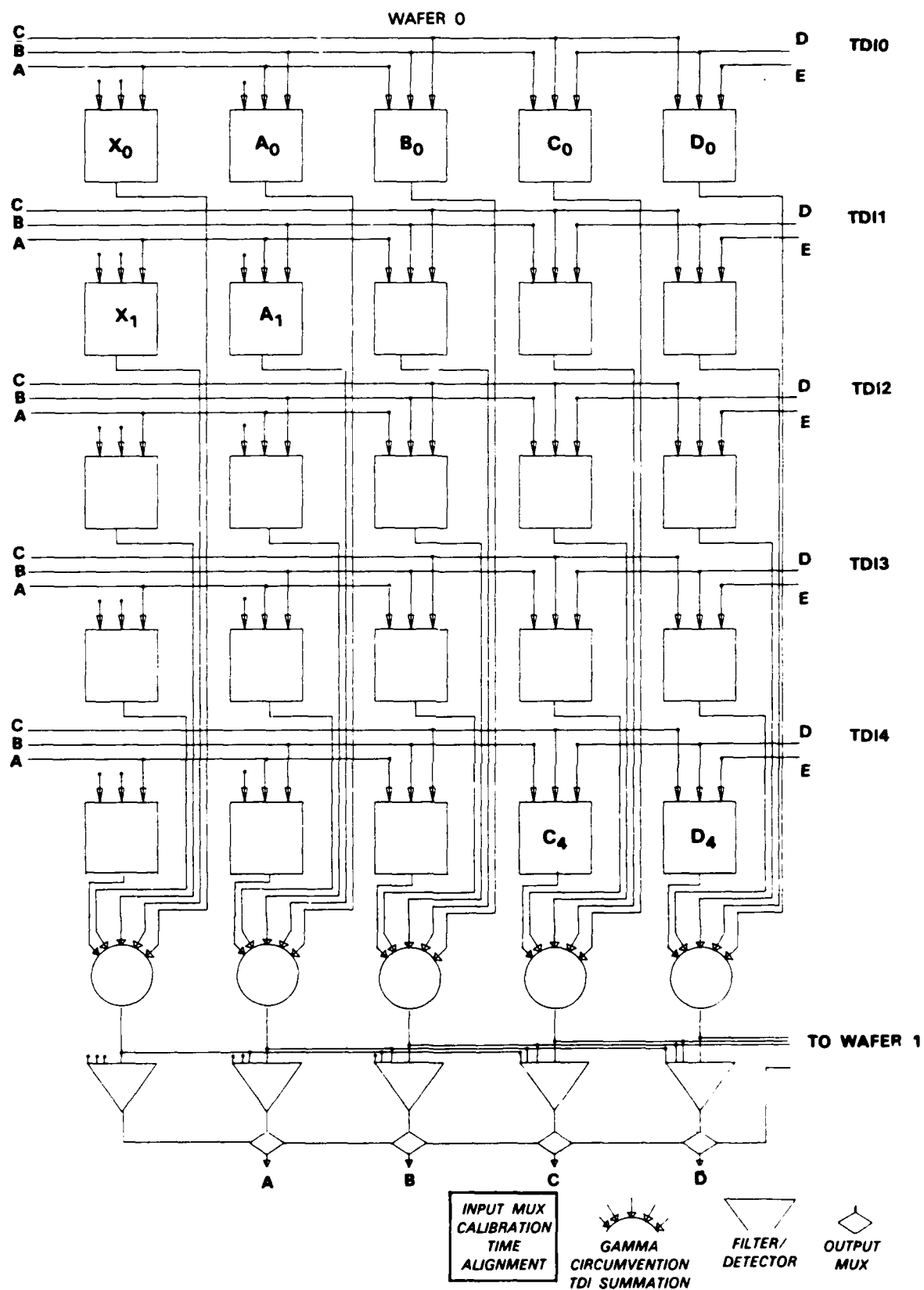
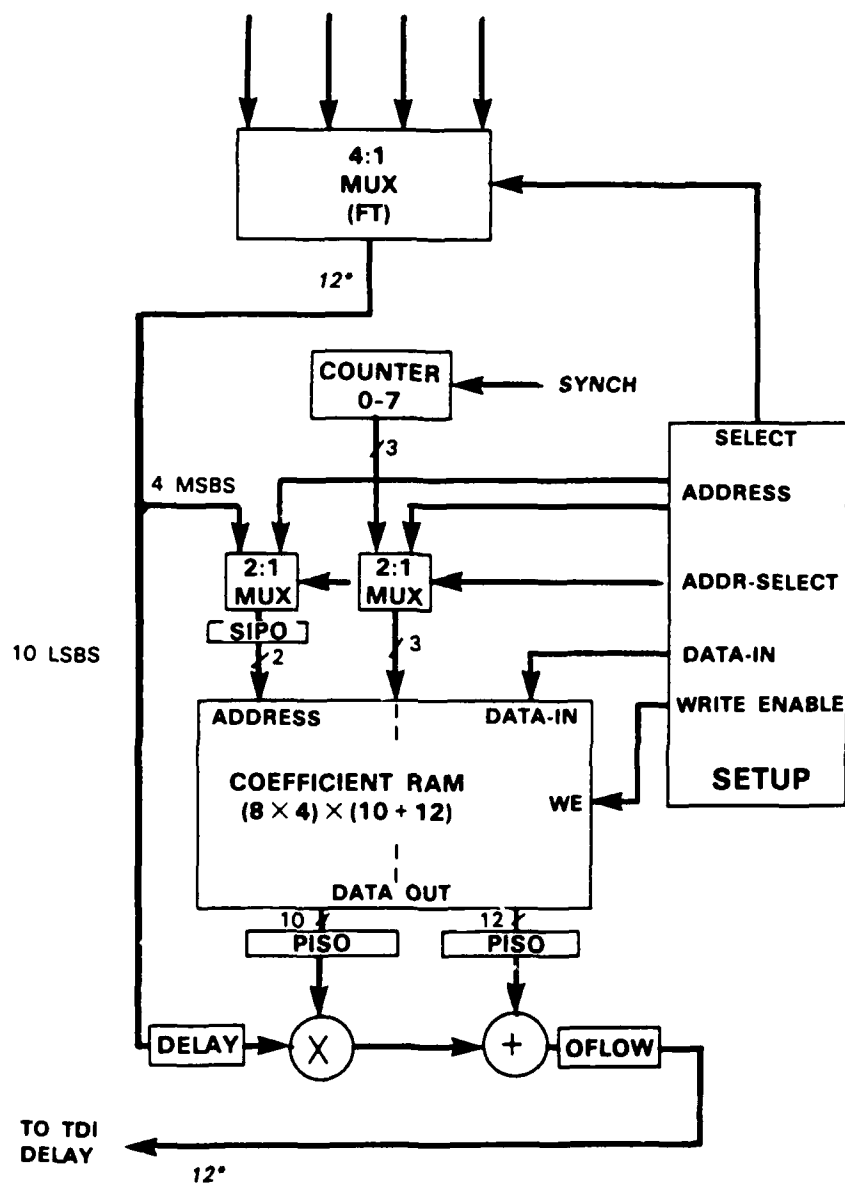


Figure 2-2. Schematic wafer layout.



* PROCESSOR IS SERIAL WITH BASIC PERIOD OF 14 BITS. THIS NOTATION REFERS TO NUMBER OF NONZERO BITS EXPECTED.

Figure 2-3. Input and calibration cell.

calibration memory contains $8 \times 4 \times 22 = 704$ bits, which is quite modest. Memories that small tend to be dominated by their address decoding logic; an increase in integration time on the part of the focal plane would permit more rows to be handled by each PE and a concomitant increase in storage efficiency.

Although the circuit is designed to implement a 4-segment linear correction, several other functions are possible using the same circuit, but with different data in the memory, notably a simple gain and offset calibration. At present, IR systems typically use either single point (offset) or 2-point (gain and offset) calibration, due to the difficulty of finding appropriate calibration standards in the infrared. This situation is unlikely to change in the near term; the requirement of 12-bit accuracy thus translates into a rather daunting requirement on the photodiode array of linearity better than 1 part in 4096.

The SETUP logic on the right of Figure 2-3 controls the downloading of the calibration coefficients from off-wafer. Note that aside from the overflow protection, no attempt is made in the on-wafer logic to impose any reasonableness criteria on the coefficients (e.g., continuity at segment boundaries). This is the responsibility of the off-wafer calibration algorithm.

Data representation throughout the processor is positive only. This convention does not result in any loss of generality. The detector element with the highest dark current will have an offset of zero in an all positive scheme. Other elements will have pedestals added to match it. The pedestal may then be compensated out at the output threshold. Note, however, that "hotter" (higher dark current) pixels still effectively compress the available dynamic range of the processor. Allocating 1 of the 12 bits to a sign cuts the range by a factor of 2, but with a detector uniform to ± 5 percent, the largest pedestal is 410 out of 4096, giving the edge to the all-positive approach.

2.4 TIME ALIGNMENT

The time delay and integration process requires that earlier columns be delayed so that they can be processed along with later ones. In the strawman design under consideration, the unit TDI delay is $28 \mu\text{s}$. Since the last stage need not be delayed, time alignment consists of delaying each set of 32 detector inputs by 0, 28, 56, 84, or $112 \mu\text{s}$, respectively. (Recall that there are 8 detector rows per PE and each dwell is oversampled by a factor of 4.) Logically, the delay stages may be thought of as delay lines. However, implementing delay lines in CMOS is undesirable because of the large switching currents. Instead, the delays are implemented as circular buffers, in which only the address pointers are incremented while the data remain in place. A single delay stage, which is a restructurable cell in the design, is shown in Figure 2-4. Incoming data arrive in bit-serial format, are converted to parallel with a serial-in-parallel-out (SIPO) converter and are stored in a 32×12 static RAM. The read and write addresses for this memory are controlled by a counter. Since the delay is 32 words, word $n + 32$ always overwrites the location that word n was just read from. Multiple delays are implemented by daisy chaining this 32-word delay cell. The small capacity memory cell is not area-efficient by itself, but the efficiency of not constructing 4 different size memories more than compensates.

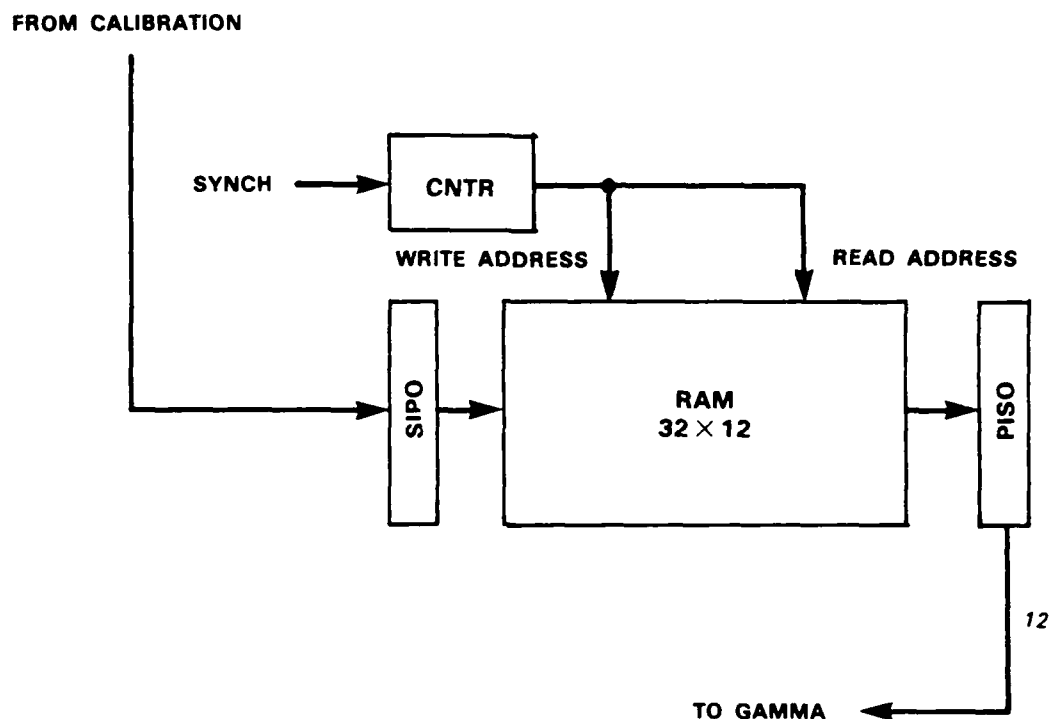


Figure 2-4. Time delay cell.

2.5 GAMMA CIRCUMVENTION AND TDI SUMMATION

The purpose of this cell is twofold: reject detector element signals which have been contaminated by γ events and then average the remaining TDI elements together. Before turning to the implementation on the PFPP, we will give a brief introduction to gamma circumvention (in order to motivate it) and an alternative approach.

What is being circumvented in gamma circumvention is noise produced not directly by γ s, but by electrons produced by the interaction of γ radiation with matter in the vicinity of the detector array. The interaction of γ radiation with matter takes place through three main mechanisms:

- (1) Photoelectric effect
- (2) Scattering on free electrons
- (3) Pair production

At the energies associated with nuclear-produced radiation, items (1) and (2) are the dominant mechanisms. (See, for example, [2] section 2-9 for a discussion of the physics.) The resultant electrons are charge carriers which produce effects in the detector similar to those produced by

IR-photon-induced carriers. They produce an energy spectrum with a long exponential falloff ("Landau tail") characteristic of the passage of ionizing radiation through matter.

2.5.1 Algorithm

The algorithm chosen is a variant of the Spike Adaptive (SATDI) type[15,16]. Many variants of SATDI exist, but all rely on the basic idea that detector response within a TDI set should be the same within some noise variation. Any sample outside some statistically determined limit is then assumed to be contaminated with a " γ " pulse, and is eliminated. A common approach (assuming unipolar spikes) is to use a lowest-of-N algorithm, in which the lowest TDI sample is considered to be the one most likely free of contamination. This algorithm is easy to implement in digital logic. Because of its theoretical attractiveness, however, the approach taken here is to model the data as a Poisson random variable with mean λ and standard deviation $\sqrt{\lambda}$. The estimated parameter $\hat{\lambda}$ is formed by summing the 5 TDI samples and scaling by $\frac{1}{5}$. The threshold is then formed as

$$\hat{\lambda} + k\sqrt{\hat{\lambda}}$$

where k is the number of standard deviations used.[†] A TDI sample which exceeds the threshold is considered contaminated and excluded.

2.5.2 Alternate Approach

The thrust of all SATDI approaches is to consider the γ -contaminated samples to be bad data, eliminate them, and proceed with processing on the remaining data. The SATDI approach has two disadvantages:

- (1) The signal-to-noise ratio is degraded, for the discarded samples no longer contribute to the \sqrt{N} SNR gain.
- (2) The output becomes biased, as the γ threshold eliminates samples with large positive random variation.

An alternate approach is to perform a maximum likelihood detection algorithm on the signal in the presence of γ noise. This approach is feasible if a parametric form of the γ noise is assumed, and is explored in more detail in Appendix A.

2.5.3 Threshold Generation

Figure 2-5 shows the schematic for the gamma circumvention and TDI summation cell. It is

[†] Note that Poisson statistics apply in this form only to the raw photodetection process. If the detector output has been scaled down by some factor s at the input stage, then the standard deviation becomes $\sqrt{s\lambda} = \sqrt{s}\sqrt{\lambda}$. Thus, the threshold factor k must effectively be rescaled by \sqrt{s} .



broken into two logical sections: the upper part generates the SATDI threshold, while the lower part compares each TDI sample with the threshold and averages the accepted samples.

Threshold generation in SATDI is a reasonably heuristic affair; consequently there is no requirement of extreme precision in the threshold generation circuit. The $\frac{1}{5}$ circuit is approximated by

- (1) Summing the 5 inputs
- (2) Rounding and shifting right 2 bits, leaving a 13 significant digit sum
- (3) Multiplying the sum by $\frac{4}{5}$, approximated to 6 bits as $0.110011_2 = 0.796875_{10}$ for a 0.4 percent error.

See Appendix B for a further discussion of this approach.

To save space, the square root is calculated using a 256×6 ROM. As shown in Figure 2-5, the method is a 2-range lookup table. The 12-bit input data are shifted left 4 bits if the data item is less than 256, and the resulting 8 MSBs are then used to address the table. This shift maps the ranges 0 to 255 and 256 to 4095 into a single 256 element table. The output of the table is compensated by shifting right 2 places if the input is shifted left. The dual-range lookup yields a maximum difference of 1 from a true integerized square root over the range 0 to 4095.

The output of the square root table, which represents an estimate of the standard error, is then multiplied by a 5-bit γ constant and the product is added back into the delayed average to form the SATDI threshold. The multiplier is arranged so that the output is scaled by $\frac{1}{8}$. Thus, the γ constant is effectively in the form xx.xxx, allowing a range of 0 to 3.875 in steps of 0.125.

2.5.4 Comparator and TDI Summation

The output of the SATDI generation circuit is fanned out and compared with the delayed TDI set in parallel. Those elements which are under threshold are passed through to a summer. The output of the comparator is also passed to a circuit which generates a multiplier for scaling the summer output. The multiplier is $4/N$ rather than $1/N$ because the summer has prerounded and right-shifted the sum bits by 2, in order to guarantee that the maximum number of nonzero bits is 13. Proceeding in this manner, which is advantageous from a hardware point of view, can cause an error in the least significant bit. The effect is not significant except when the signal and background are both small. Appendix B contains a more detailed discussion.

2.6 MATCHED FILTER AND DETECTOR

The matched filter is a separable 4×4 digital filter. Being separable means that the filter may be constructed as the convolution of a 4-tap filter oriented vertically with another 4-tap filter oriented horizontally, resulting in a savings in the amount of computation required. The 4-tap

horizontal filter is matched to the $4\times$ in-scan oversampling. The 4-tap vertical filter assumes that the cross-scan resolution is made comparable to the in-scan resolution by using rectangular pixels.

The matched filter and detector cell is shown in Figure 2-6. A far more detailed hardware description is given in [3]. Much of the complication of the interconnect stems from the fact that the cross-scan filter requires data from adjacent detector rows and hence adjacent PEs. The most naïve design would require data from both nearest neighbors; the current implementation offsets the cross-scan filter so that a PE only requires data from the previous PE (Figure 2-7). The output of the PE can then be shifted up to compensate for this offset (ignoring edge effects).

The 8×12 delays are implemented as true shift registers in this design, resulting in a significant current draw. A more capable PE would require longer delays and these could also be implemented as circular buffers like the time alignment memories.

A simple threshold detector is attached to the output of the horizontal convolution. The threshold is loaded from off-wafer, so that it can be adjusted during operation of the PFPP as a means of controlling the overall false alarm rate. The comparator is implemented as a combinatorial full adder, and the full signed difference is sent off-wafer to the ODP.

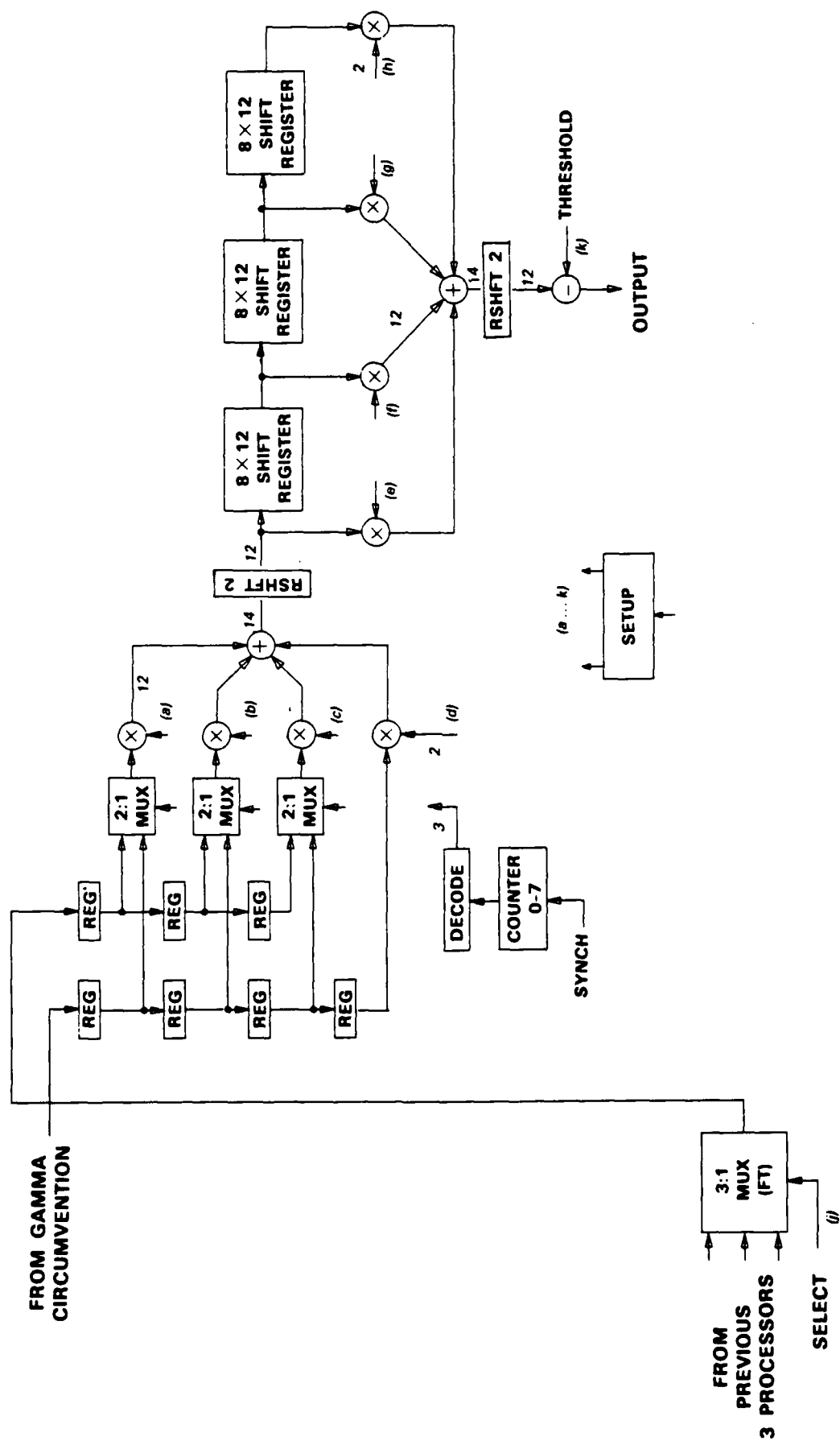


Figure 2-6. Matched filter and detection cell.

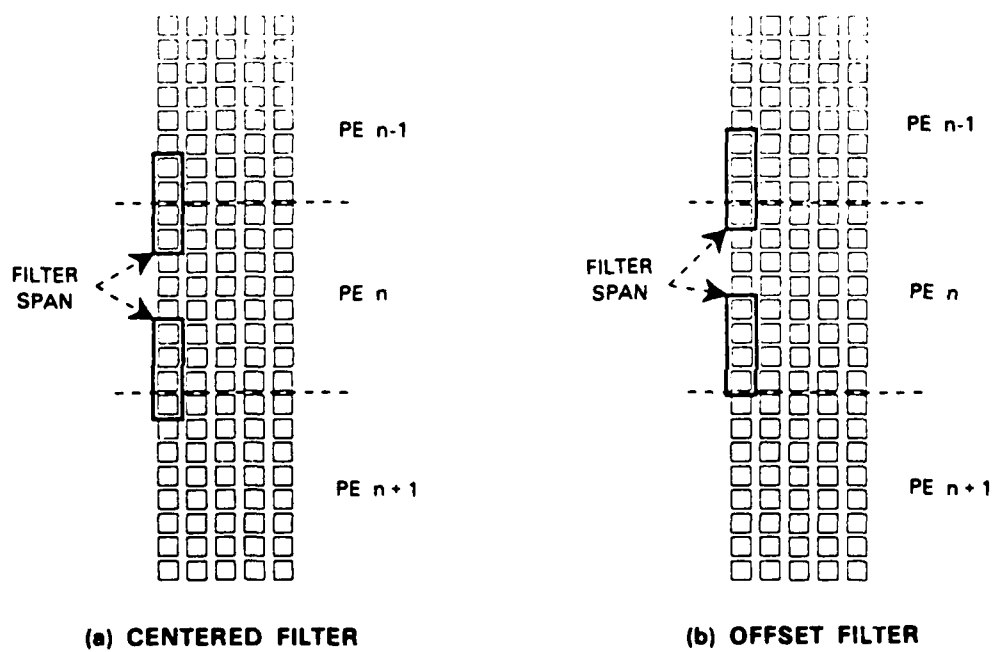


Figure 2-7. FIR filters.

3. IMPLEMENTATION ISSUES

3.1 AREA ESTIMATES

In the course of the conceptual design of the wafer, area estimates were done using relatively crude estimates of the circuit elements needed in the design. The area required for certain circuit components was estimated as shown in Table 3-1. The estimates for static RAM and shift registers

TABLE 3-1.
Component Area Estimates

Component	Area (mm ²)
Static RAM, per bit	0.0063
Shift register (static), per bit	0.0081
Serial multiplier, per bit	0.1200
Serial adder, per bit	0.0580
Tristate register, per bit	0.0225

were from designs being developed by Group 23 at Lincoln Laboratory. The memory figure assumed a cell size of $40\lambda \times 50\lambda$, and amortized the area required for read/write and address select over the per-bit figure. This area is nonnegligible for small memories as are used in this design. The remaining estimates were from MOSIS scalable designs with $\lambda = 1.5$ (for 3- μ m technology).

3.2 SERIAL VERSUS PARALLEL ARITHMETIC

Using these figures, area estimates for both a serial and a parallel arithmetic processor were developed. For equal clock rates, the parallel processor will have 12 times the capability of the serial one.* One might naïvely expect each serial processor to be $\frac{1}{12}$ of an equivalent parallel processor in area (see [12], p.20). This is not the case for a number of reasons:

- (1) Extra accumulator registers have to be provided for the multipliers.
- (2) Extra shift registers have to be provided for increased latency at choke points where all bits are required (e.g., calibration, gamma circumvention).

* For ease in supplying input data, the prototype parallel processor was sized for only 16 detector rows rather than 96; it was designed to run in burst mode, with a low duty cycle.

- (3) Calibration and TDI delay memories become less dense as their size is reduced to serve fewer detectors.
- (4) Since serial adders are pipelined, intermediate word growth that appears when summation is followed by division, e.g., TDI summation, has to be accommodated by padding out the bit stream with extra zeros.

Figures 3-1 and 3-2 are graphical representations of the area estimates for the PFPP using parallel and serial arithmetic. Inset into them, in turn, are Tables 3-2 and 3-3 which present the data numerically and serve as the figure keys. In the figures, space allocated to a PE is represented by the horizontal chaindash bars. Within the bars, shaded boxes represent area allocated to circuit elements; white space around the boxes is reserved for interconnect. The crosshatched areas at the left and right are input and output buses.

The figures graphically illustrate the smaller granularity of the bit-serial architecture. Due to the significantly larger size of a parallel-arithmetic PE, and the width of the buses, only 6 (optimistically) would fit on a 3-inch wafer. The sparing strategy was to have 3 working PEs (2 active and 1 spare) per wafer. Eighteen serial processors were calculated to fit in the 50-mm square contained within a 3-inch wafer. Since only 5 are needed in the prototype system, this approach would permit a working processor even if early yields in the SOI process were relatively low.

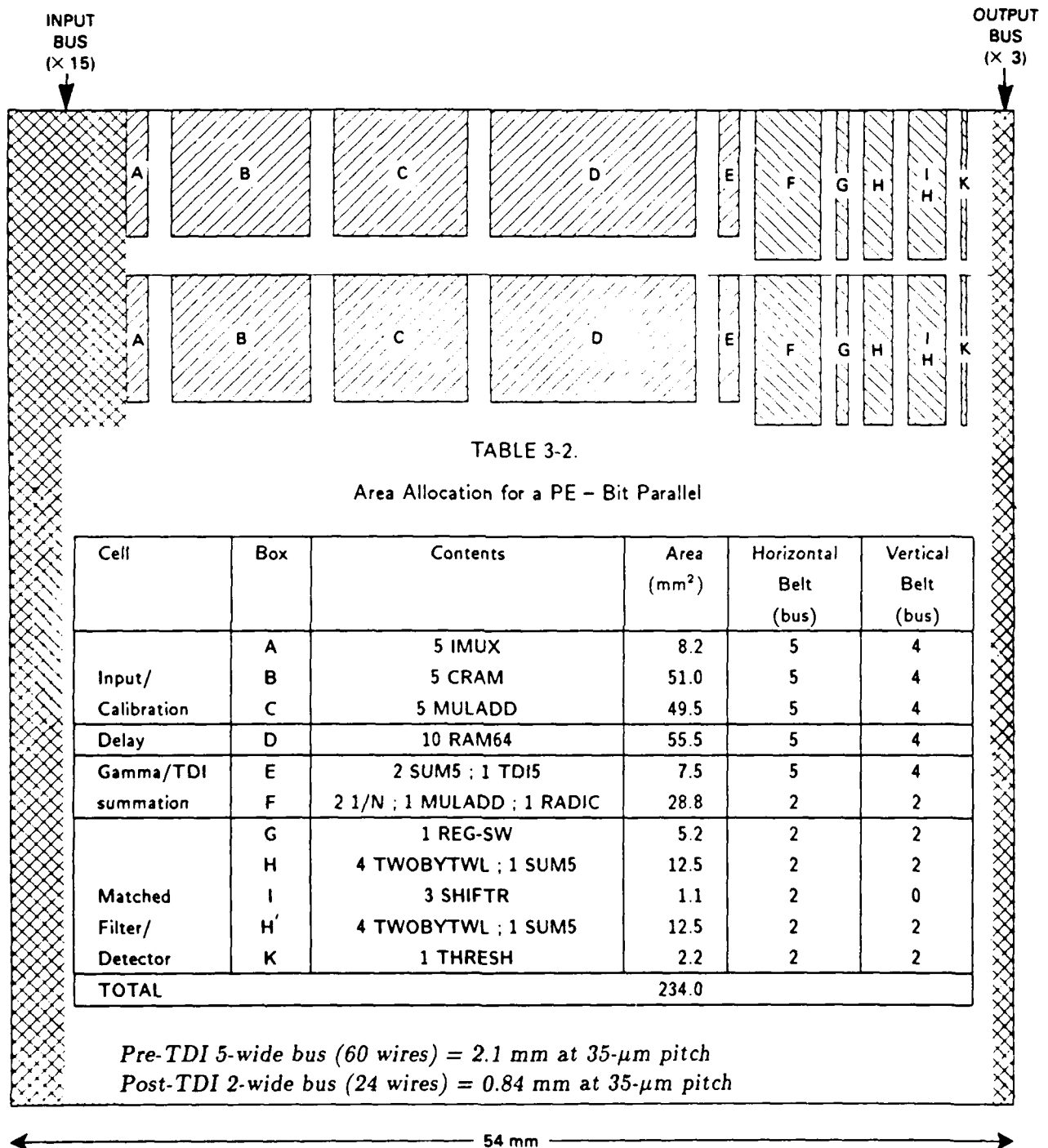


Figure 3-1. Schematic area allocation on PFPP wafer - bit parallel.

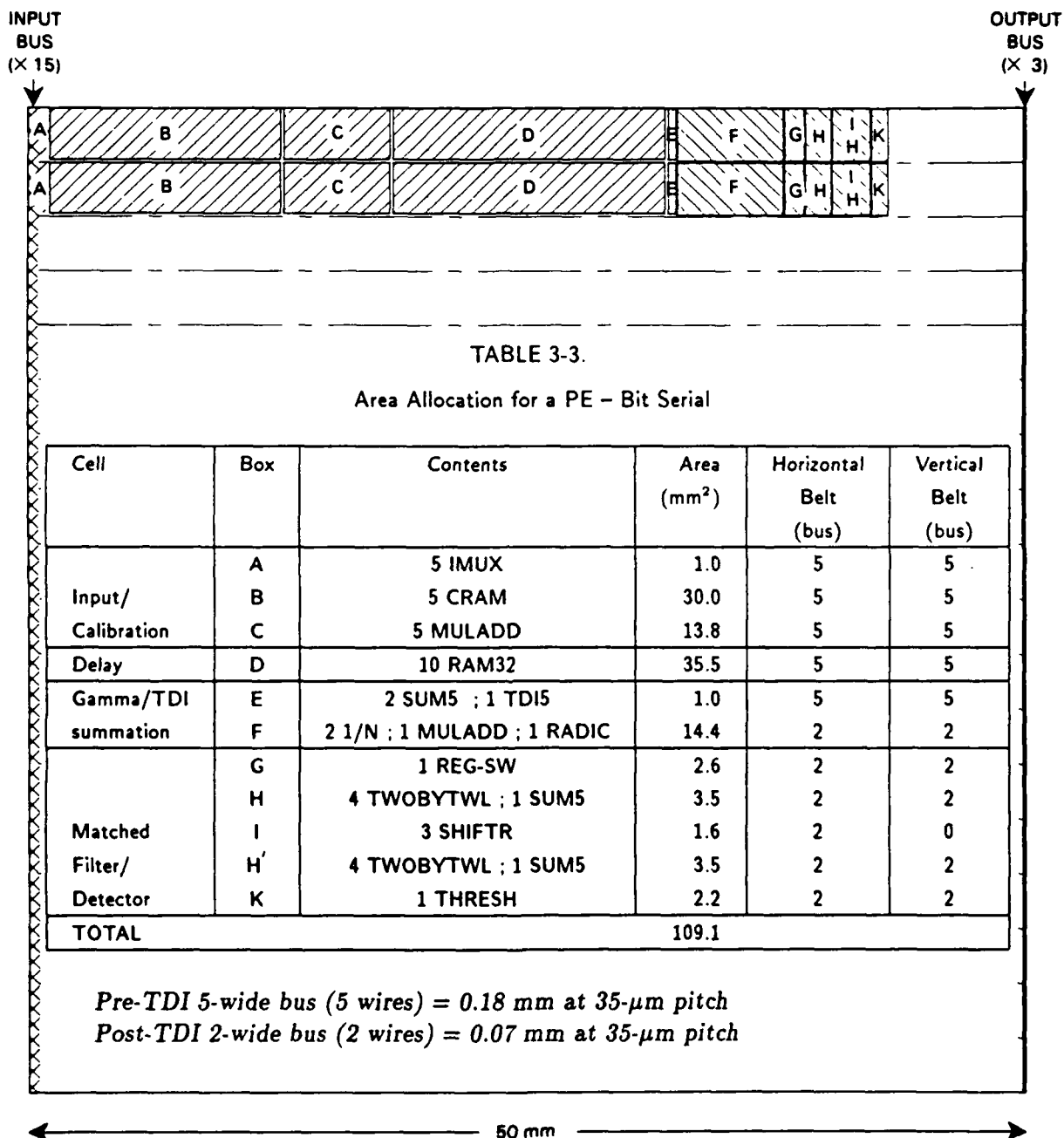


Figure 3-2. Schematic area allocation on PFPP wafer - bit serial.

3.3 PREDICTED VERSUS ACTUAL AREA

Another interesting comparison can be made between the estimates given in Table 3-3, which were generated roughly a year before the present report was written, and the actual area taken by the cells. As of this writing, two of the four cells (TDI delay and matched filter/detector) have been designed by Group 23 and received back from MOSIS. A third (input and calibration) is in final layout and its size can be estimated with confidence. The fourth (gamma circumvention/TDI summation) is well along, and its area can be estimated with reasonable accuracy. This comparison is made in Table 3-4.

TABLE 3-4.
Estimated versus Actual Cell Area

Cell	Estimated (mm ²)	Actual (mm ²)	Difference (%)
Input/calibration	9.0	≈10	≈11
TDI delay	3.5	3.4	-3
Gamma/TDI summation	15.4	≈17.5	≈14
Matched filter/detector	13.4	10.0	-25

Agreement between the rough calculations and as laid-out areas is remarkably good. At the time of the initial calculation (16 June 1987), Group 23 had a reasonable idea of what its small static RAM would look like; hence the input/calibration and delay estimates are much closer than the other two cells. Early memory estimates, based on large commercial RAMS, had tended to be much more optimistic. Since the errors on gamma/TDI summation and matched filter/detector roughly cancel, it is reasonably certain that the goal of laying down 18 PEs on a 50-mm square could be met.

Early results with wafer scale circuits implemented in the Lincoln Laboratory zone melt refined (ZMR) SOI technology (see [10] for a review), however, indicate that defects in ZMR wafers tend to occur preferentially at the edge of the wafer. Therefore, the preliminary FPP is being designed to fit into a 40-mm square, allowing a 5-mm buffer on all edges: $(\frac{40}{50})^2 \times 18 = 11.5$; because of inefficiencies in packing, probably about 10 PEs will fit in this smaller area.

4. RELIABILITY ESTIMATES

One of the most stressing demands on a focal plane processor is the requirement of reliability in a space environment. The FPP is designed for a nominal five-year lifetime. Modern integrated circuit design results in highly reliable circuits. The extremely large number of circuit elements, however ($\approx 12,000$ transistors in the matched filter/detector cell alone), results in a rather small total probability for a system working perfectly for five years.

It is vital to design reliability in from the beginning in order to have any realistic hope of achieving mission requirements. On the other hand, precise reliability measurements are obviously lacking in any new design, and more so than usual in rad-hard wafer-scale technology. As discussed previously, the approach taken for the PFPP is to utilize a redundant network of processor elements. In order to evaluate this approach quantitatively, the reliability of an individual PE must be estimated; if the PE is too complex, the survival rate will be too small. In this case, sparing must be provided at a lower level, or a large number of spare PEs must be allocated. Thus, some sort of estimate of PE reliability must be found in spite of the novelty of the technology. A certain amount of sloppiness in the estimates must be tolerated, and the sensitivity of the overall PFPP reliability to this uncertainty must be at least estimated.

4.1 MIL-HDBK-217E

Recognizing the problem, DoD has issued MIL-HDBK-217E, *Reliability Prediction of Electronic Equipment* [4]. This handbook presents failure models of electronic components and systems, as well as constants for evaluating the models based on experience to date.

Section 5.1.2 of [4] presents a failure rate prediction model for monolithic microelectronic devices. This model is:

$$\lambda_P = \pi_Q \cdot (C_1 \pi_T \pi_V + C_2 \pi_E) \cdot \pi_L$$

where

λ_P	is	the predicted device failure rate in failures/ 10^6 hours,
π_Q	is	the quality factor,
C_1	is	a circuit complexity factor, depending on transistor count and technology,
π_T	is	the temperature acceleration factor,
π_V	is	the voltage stress derating factor,
C_2	is	a package complexity factor,
π_E	is	the application environment factor, and
π_L	is	the device learning factor.

For the purposes of our discussion, let us consider this as

$$\lambda_P = \pi_Q \pi_L \cdot \underbrace{(C_1 \pi_T \pi_V)}_{\text{Term 1}} + \underbrace{C_2 \pi_E}_{\text{Term 2}} \quad (4.1)$$

4.2 PROCESSOR ELEMENT RELIABILITY ESTIMATE

Term 1 applies to failures of individual PEs, and hence will be reduced by PE sparing; Term 2 applies to packaging failures. Since the pinouts are not redundant in the current design, Term 2 will not be affected by our sparing strategy – failure of a pin will reduce the capability of the wafer to perform its mission. Clearly there is not a lot of field experience with wafer-scale SOI. However, in order to proceed with quantitative analysis of a PE, we must evaluate the various factors of Term 1 as best we can. (Note that references to Table 5.1.... in this and Section 4.3 are to tables in [4], not this report.)

C_1 : As might be expected, MIL-HDBK-217 has no data directly applicable to wafer-scale devices. Hence, values for C_1 must be extrapolated from data it presents. Of the devices that might be applicable to the FPP case, Section 5.2.1 presents data for shift registers, static RAMs, and microprocessors in CMOS. The approach taken here is to simply calculate the C_1 for each, and then sum them. (Adding probabilities of failure is equivalent to multiplying probabilities of success as long as $P_F \ll 1$) We have

Device	C_1
Shift register (<1000 gates)	0.02
Static RAM (<16 K)	0.10
Microprocessor (16 bit)	0.06
Total	0.18

π_T : This factor depends on the technology and the worst-case junction temperature. For the *space flight* environment, worst-case case temperature is specified as 45 °C. The rise over case temperature is difficult to estimate with any precision; the PFPP wafer may dissipate ≈ 5 W (the Lincoln Laboratory fast Fourier transform wafer dissipates ≈ 3 W at 16 MHz)[6]. This heat will not be produced uniformly over the surface of the wafer, however. A wild guess for the worst-case junction temperature rise is $T_J = 15$ °C over the case temperature, or 60 °C. This choice yields $\pi_T = 0.95$ from Table 5.1.2.7-8.

π_V : This is 1.0 from Table 5.1.2.7-14.

4.3 FPP RELIABILITY ESTIMATE

To complete the analysis of the wafer, we will first evaluate Term 2 of equation 4.1 to gain an estimate of wafer reliability without sparing; then we will add in the sparing combinatorics.

C_2 : The package complexity factor is given in Table 5.1.2.7-16 as

$$C_2 = 3.0 \times 10^{-5} N_P^{1.82}$$

for hermetic flatpacks. This equation is only valid up to $N_P = 24$ pins. However, blithely extrapolating to 40 pins for the prototype FPP, we obtain $C_2 = 0.188$. Note that for a 6-inch wafer of ≈ 500 pins, this equation yields $C_2 = 2.5$, suggesting that redundant pinouts (or much improved packaging) will be an important part of the design strategy for the full-up FPP.

π_E : The space flight environment S_F is relatively benign, and from Table 5.1.2.7-3 $\pi_E = 0.9$.

π_L : The learning factor π_L is taken to be 10 in the case of a new device in initial production and/or a new and unproven technology.

π_Q : The quality factor π_Q is keyed to the military classification system established in MIL-STD-883. It is given as

Class	π_Q
S	0.25
S-1	0.75
B	1.0

Since the FPP would clearly not be listed on QPL-38510, I have considered it to be Class S-1, and assigned $\pi_Q = 0.75$.

For small numbers of PEs, there are roughly 8 pins per PE (5 TDI inputs, 1 output, 1 prior PE reference, and 1 control). Rewriting equation 4.1 in terms of N_{PE} , the number of PEs, and evaluating constants, we obtain for a wafer without sparing

$$\lambda_P = 7.5 \cdot \underbrace{[0.17 N_{PE}]}_{\text{Term 1}} + \underbrace{2.7 \times 10^{-5} (8 N_{PE})^{1.82}}_{\text{Term 2}} \quad (4.2)$$

For $N_{PE} = 8$, Equation 4.2 evaluates to Term 1 = 1.36 and Term 2 = 0.05, confirming our intuition that (for relatively simple packaging) PE sparing is most important. Wafer reliability over 5 years is evaluated using Equation 4.2 in Figure 4-1(a) for various numbers of PEs on a single wafer without sparing. As can be seen, although the probability of survival for a single PE $P_0 = 0.946$, for the 8 PEs required to do the job, predicted reliability $P_0^8 = 0.64$, which is rather poor.

Figure 4-1(b) shows the effect of sparing for a 2-wafer set. System reliability rises dramatically as the first 2 spares are introduced, then levels off and begins to fall slightly, so that additional spares are not helpful. This effect is due to Term 2, the projected packaging failure rate, since the PE sparing term alone continues to rise to 1. Hence, redundant pinouts will have to be introduced to raise system reliability above the ~ 97 percent level in this failure model.

4.4 INFLUENCE OF PAIRWISE INTERACTIONS

A further question which arises in the context of FPP reliability calculations is the effect of correlated failures. Section 4.3 assumes that the individual PE failure rates computed according to MIL-HDBK-217E are statistically independent, and combines them accordingly.

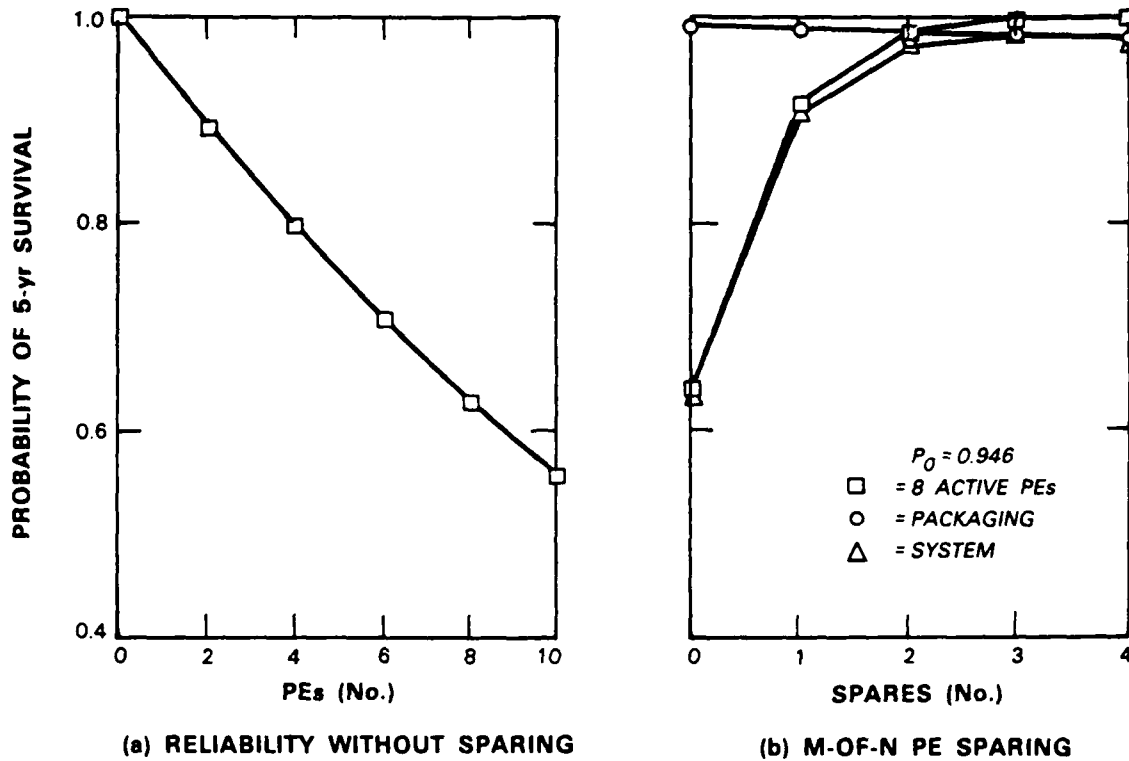


Figure 4-1. Processor reliability.

However, it is possible that the failure of one PE might somehow stress its neighbors (e.g., by dragging down bus voltage). In this case, failure of one PE would increase the probability that its neighbors would fail, so that $P_{fail}[PE_i|PE_{i\pm 1}] > P_{fail}[PE_i]$.

The approach taken here is to assume that the failure probability calculated from MIL-HDBK-217E is that of an isolated processor; nearest-neighbor interaction is then added by multiplying by a tri-diagonal interaction matrix containing the nearest-neighbor interaction C_{NN} (This approach assumes $P_{fail} \ll 1$):

$$\begin{bmatrix} P_{fail_1} \\ \vdots \\ P_{fail_8} \end{bmatrix} = [P_{fail_1} \cdots P_{fail_8}] \begin{bmatrix} 1 & C_{NN} & 0 & 0 & \cdots \\ C_{NN} & 1 & C_{NN} & 0 & \cdots \\ 0 & C_{NN} & 1 & C_{NN} & \cdots \\ & & & \ddots & \\ \cdots & 0 & C_{NN} & 1 \end{bmatrix}.$$

Not all P_{fail_i} are the same using this approach, and the sparing calculation is modified to incorporate this fact.

Figure 4-2 presents the results of the a calculation of the 5-year probability of success for various C_{NN} as a function of the number of spares, using the PE failure rate calculated from MIL-HDBK-217E. For $C_{NN} = 0.2$, 2 spares are adequate to keep the overall probability of success above 0.95.

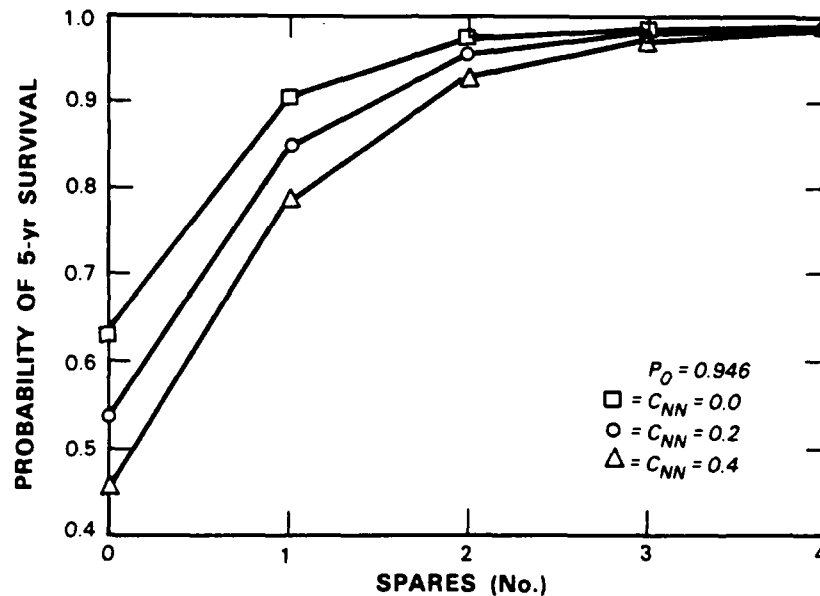


Figure 4-2. Processor reliability with nearest-neighbor interaction.

Since the rate calculations from [4] are far from exact, it is worthwhile looking at what sort of margins the sparing strategy provides. Thus, we conclude this discussion by examining the reliability of the 8-processor PFPP system as a function of the probability of survival of the individual PEs. Figure 4-3 shows the overall probability of survival as a function of a single PE's probability of survival, for the case of $C_{NN} = 0$ and $C_{NN} = 0.2$. For the no-interaction case with 2 spares, the system probability of 5-year survival stays above 0.9 as long as the single-PE probability is also above 0.9. With $C_{NN} = 0.2$, this threshold is increased to ~ 0.925 . Increasing the number of spares to 4 gives a system probability of survival > 0.9 as long as an individual PEs probability of survival is > 0.8 for the no-interaction case.

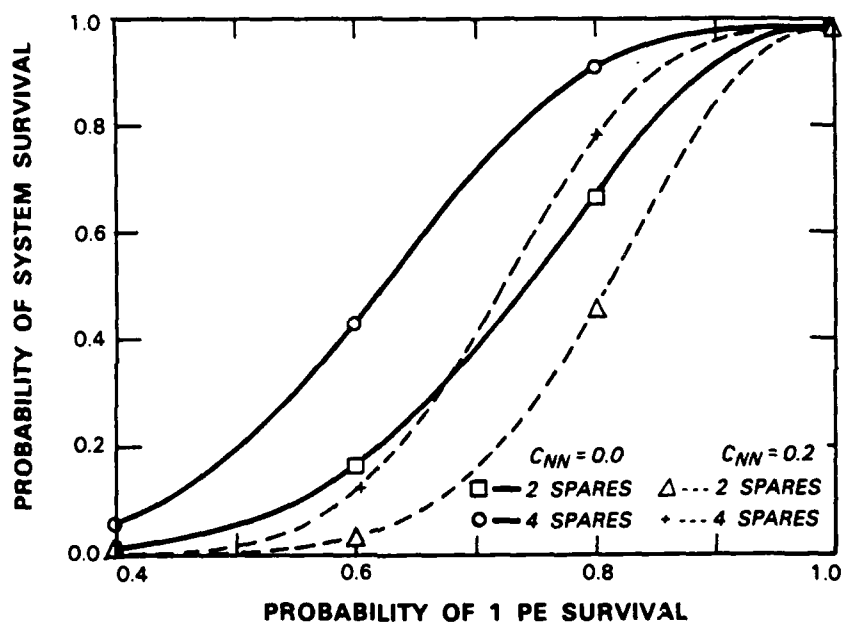


Figure 4-3. Effects of PE reliability.

103581-12

5. CONCLUSION

We have described the design of a prototype wafer-scale focal plane processor. This design represents an evolution from previous work in monolithic VLSI, as it has four different cells in each PE, rather than a relatively uniform array of cells as in earlier work[7,8]. The design incorporates fault-tolerant technology in order to achieve a five-year lifetime with predicted 97 percent confidence.

Wafer-scale technology represents an important advance in focal plane processors for space-based applications. Current processors in aircraft-based applications, with only medium levels of integration, occupy several cubic feet of volume and dissipate several thousand watts[5]. By comparison, a processor based on wafer-scale technology would be at least an order of magnitude smaller in both parameters.

APPENDIX A

MAXIMUM LIKELIHOOD APPROACH TO GAMMA CIRCUMVENTION

Since the early 1970's, there has been a great deal of effort expended on increasing the immunity of infrared sensor systems to the effects of γ radiation. The most successful approach by far has been to harden the focal plane. Developments in detector technology have resulted in an enormous decrease in detector volume, which has reduced the detector cross section by several orders of magnitude. The consensus is that the easy gains have been made, so that absent any breakthroughs such as intrinsic event discrimination, the next order-of-magnitude increase will be a lot harder than the previous four.

At the same time, electronic γ circumvention has made great strides. Given the difficulty of pushing the state of the art in detectors and the computational resources expended on circumvention, it is worthwhile to reexamine the field.

The most successful approach to γ circumvention to date is a two-stage algorithm, the spike adaptive time delay and integration (SATDI) method pioneered by Boeing. This heuristic approach makes no assumptions about γ -induced noise except that it is an additive corruption of the true signal (assuming unipolar γ spikes, as we will do throughout). The first stage of the method utilizes all the TDI signals corresponding to a given point in space to set an upper bound on a reasonable signal. Signals above this bound are assumed to be corrupted by γ spikes, and are discarded. In the second stage, detection proceeds normally on the cleaned-up sample.

A.1 MAXIMUM LIKELIHOOD MODEL

If, however, one is willing to make assumptions about the form of the γ -induced noise distribution, it is possible to design an optimal detector for a given signal in the presence of this noise utilizing classical maximum likelihood detection theory. The rest of this appendix reports on such a detector, utilizing the formalism developed in H. L. Van Trees[14], Chapter 2.

The parametric form chosen is an exponential distribution, $\lambda e^{-\lambda r}$ where r is the received signal, which is a reasonable approximation to the observed γ spectrum in IBC detectors.[13] The noise model is thus the sum of (a) Gaussian background noise with variance σ^2 and mean μ , (b) exponential γ noise with mean $1/\lambda$ and a probability of occurrence in a given sample f . That is, if n_T is the total noise,

$$n_T = n_B + n_\gamma \quad (A.1)$$

$$p_{n_B}(n_B) = N(\mu, \sigma) \quad (A.2)$$

$$p_{n_\gamma}(n_\gamma) = f\lambda e^{-\lambda r} + (1-f)\delta(0) \quad (\text{ignore multiple hits}) \quad (A.3)$$

Adding random variables is equivalent to convolving their pdf's, so (adopting the notation of Van Trees) the probability density under the null hypothesis is

$$p_{r_1|H_0} = f\lambda Q(-\alpha/\sigma) \exp\left[\frac{\lambda}{2}(\sigma^2\lambda + 2\mu - 2r)\right] + \frac{(1-f)}{\sqrt{2\pi}\sigma} e^{-(r-\mu)^2/2\sigma^2} \quad (A.4)$$

$$= \mathcal{F}(r_i; \mu, \sigma, \lambda, f) \quad (\text{A.5})$$

where

$$\alpha = r - \mu - \sigma^2 \lambda \quad (\text{A.6})$$

$$Q(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt . \quad (\text{A.7})$$

We let hypothesis H_1 correspond to the presence of a constant voltage m . This voltage simply shifts the scale, so that

$$p_{r_i|H_1} = \mathcal{F}(r_i - m; \mu, \sigma, \lambda, f) . \quad (\text{A.8})$$

Thus the likelihood ratio Λ is

$$\Lambda = \frac{\prod \mathcal{F}(r_i - m)}{\prod \mathcal{F}(r_i)} \quad (\text{A.9})$$

or, equivalently,

$$\log(\Lambda) = \sum \log(\mathcal{F}(r_i - m)) - \sum \log(\mathcal{F}(r_i)) . \quad (\text{A.10})$$

Note that the parameters f and λ may be estimated independently of σ , μ and m by monitoring a small sample of the detector array which is kept outside the field of view of the telescope.

A.2 ILLUSTRATIVE EXAMPLE AND COMPARISON WITH SATDI

In this section, we will consider a particular case corresponding to detection against a bright, i.e., nuclear-induced background using IBC detectors. The parameters are

Parameter	Value
TDI	8 stages
μ	400000 electrons
σ	632 ($= \sqrt{\mu}$)
λ	1/4000
f	0.3

The probability density function corresponding to equation A.4 is shown in Figure A-1. It is characterized by a Gaussian part, which falls off rapidly, and a long exponential tail resulting from the γ -induced corruption. In Figure A-2 the log of the likelihood ratio (Equation A.10) is plotted for 3 cases: no γ contamination, 15 percent γ s, and 30 percent γ s. The signal strength m is taken to be 1265 electrons, corresponding to twice the standard deviation of the background noise. The output signal-to-noise ratio is thus $2\sqrt{8}$ or roughly 5.7.

103581-16

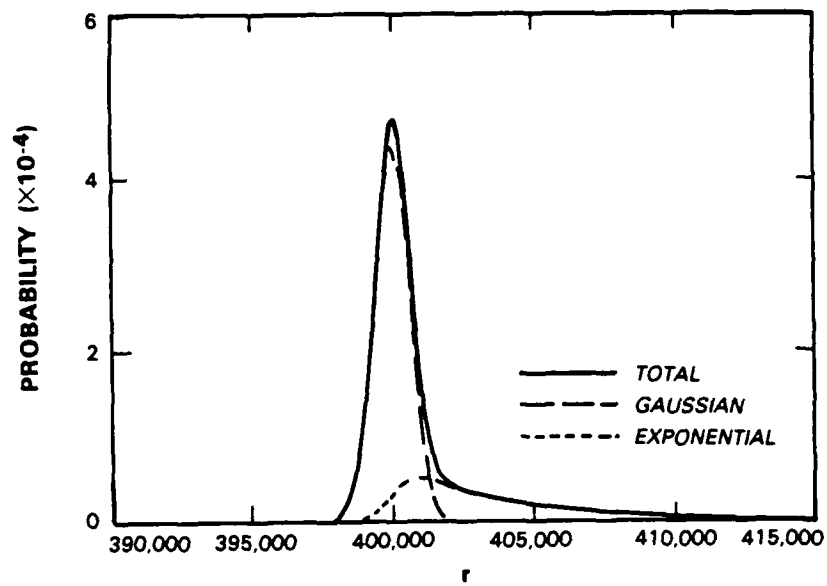


Figure A-1. Probability density function.

103581-17

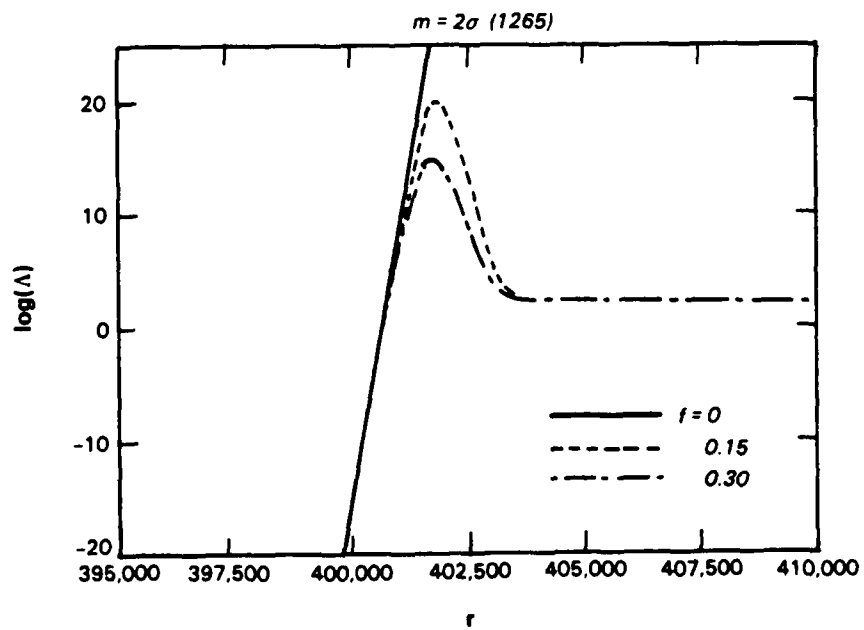


Figure A-2. Log likelihood ratio.

For the γ -free case, the log likelihood ratio is just a straight line, reflecting the classical result for a signal in the presence of Gaussian noise. Addition of γ contamination results in the ratio peaking, then falling back to an asymptote. Physically, this behavior can be thought of as reflecting the fact that high received signals are most likely due to γ contamination. The signals of interest to the detector would thus be those lying in a roughly parabolic region whose bottom is defined by the number of acceptable false alarms.

In contrast, the effect of SATDI is approximately equivalent to taking the $f = 0$ curve and cutting it off sharply at some value of r , creating a triangular region of interest.

The log likelihood ratio is a nonlinear equation and is difficult to analyze analytically, while SATDI is, by nature, heuristic. Therefore, we compared them using a Monte Carlo simulation. Events were generated using our model parameters. For the SATDI case, the algorithm used was that designed for the wafer-scale prototype focal plane processor chip. In this algorithm, σ is estimated as the square root of the average TDI signal, and the γ threshold was set at 1.2σ . The remaining signals were then averaged and compared to a detection threshold.

For the maximum likelihood case, the log likelihood of each sample was calculated, and the sum compared to a detection threshold. All other parameters, however, were known and fixed, so that this simulation does not measure the sensitivity of the method to misestimation of f , λ , and μ .

The two methods are compared in Figure A-3 by plotting the receiver operating characteristic (ROC). In an ROC plot, the probability of detection is plotted versus the probability of false alarm. The curves are approximate fits to the data points, which have significant scatter due to the limited number of Monte Carlo throws per point (20,000). A typical error bar is shown for reference. The maximum likelihood curve lies significantly above the SATDI one, indicating that either much better detection for a fixed false alarm rate, or many fewer false alarms for a fixed detection probability may be achieved.

In conclusion, maximum likelihood detection promises to give significantly better performance than SATDI within the framework of our model of γ noise. Further work must be performed to test the validity of the model, as well as to explore its performance in more realistic cases in which the signal, m , is not known a priori. In these cases, m may be estimated as is currently done with maximum likelihood detection grafted on at the end, or a maximum likelihood estimation technique may be feasible as well, resulting in a unified detection and estimation scheme.

103581-18

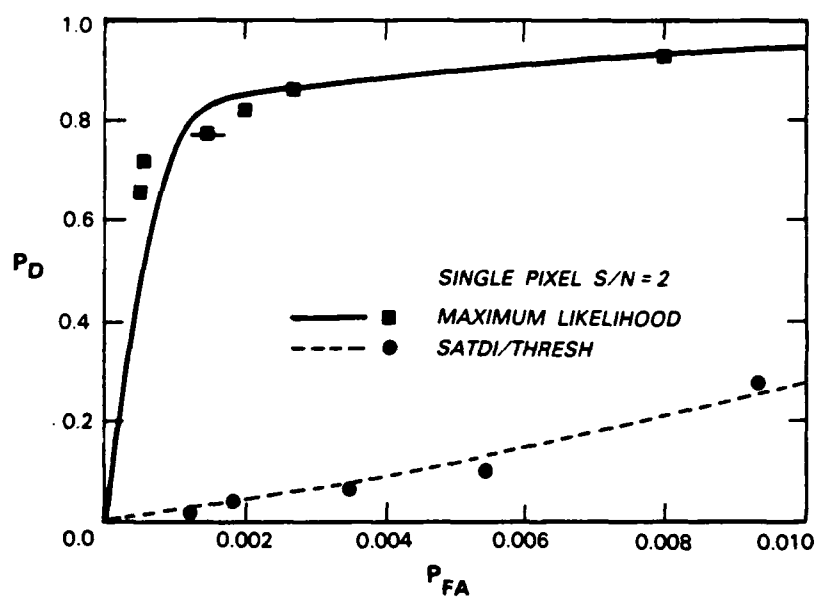


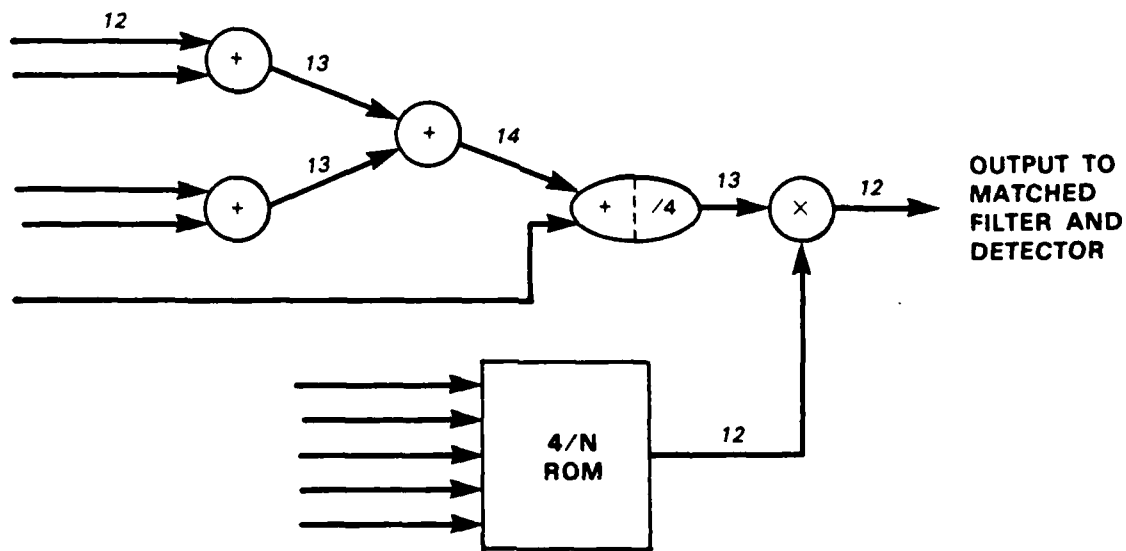
Figure A-3. Receiver operating characteristic.

APPENDIX B ROUNDOFF IN TDI SUMMATION

This Appendix describes a simulation of errors introduced by the proposed TDI summation circuit in the prototype FPP wafer.

The circuit should ideally add the N active TDI inputs ($N \leq 5$) that emerge from gamma circumvention, and then divide the result by N .

Due to implementation constraints, however, the proposed hardware implementation will look like this:



103681-13

NUMBERS ABOVE LINES = MAXIMUM NONZERO BITS

Figure B-1. Detail of proposed TDI summation circuit.

As a consequence, the final result may be off by one bit from the exact method of letting the intermediate sums cascade up to 15 bits.

As might be expected, this effect will be most severe from a percentage point of view when the background level is very low. It is also most noticeable when the γ level is low as well, since when one sample has been rejected due to γ contamination, the above method is exact.

The proposed circuit has been modeled and run through a Monte Carlo program which models the signal and background photons as a Poisson process, and incorporates γ contamination as a random exponential process according to the model of Section 2.5. The circuit was modeled for three cases:

- (1) Low background, high γ : The mean noise level was 35, the signal was 70, and the γ level was 40 percent.
- (2) Low background, low γ : As item (1), but the γ level was 0 percent.
- (3) High background, high γ : The mean noise level was 2000, the signal was 70, and the γ level was 40 percent.

The results are presented in Figures B-2 (a) and (b), and B-3 respectively. The Monte Carlos were generated assuming a 10:1 scaling at the analog to digital converters. (See footnote in Section 2.5.1.) This scale factor is used, for example, in ground-based visible data gathered by Lincoln Laboratory Group 94[11] at the Lincoln Experimental Test System in New Mexico.

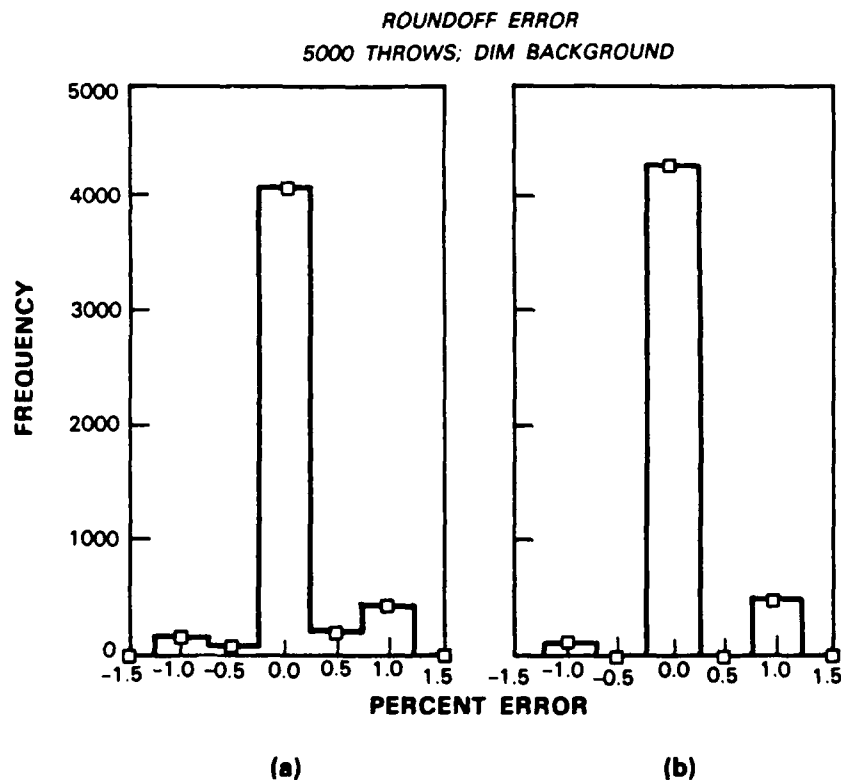


Figure B-2. Low background, (a) high gamma and (b) low gamma simulation.

As expected, the most striking error is in the low background, radiation-free case, where 14 percent of the samples will have errors of ± 1 percent (Figure B-2(b)). The forward bias is due to the Monte Carlo program's practice of always rounding $\frac{1}{2}$ up. A randomized rounding rule will equalize it.

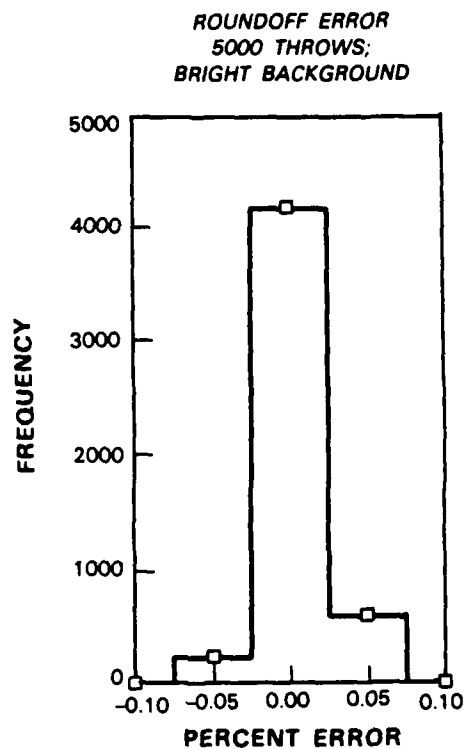


Figure B-3. High background, high gamma simulation.

This behavior is not a serious problem in a prototype processor, but it would have to be addressed in a later, full-up system in which (a) low background observations might be an important part of the mission, or (b) the number of TDI stages is increased.

REFERENCES

1. R. W. Hendrick, Jr., private communication, (1986).
2. E. Segrè, *Nuclei and Particles*, New York: Benjamin (1964).
3. C. E. Woodward, private communication (1988).
4. MIL-HDBK-217E, *Reliability Prediction of Electronic Equipment*, Washington, D.C.: U.S. Department of Defense (1986).
5. L. W. Pinkley, *Sensor Signal Processing Overview*, Huntsville, AL: Nichols Research Corporation (1987).
6. F. M. Rhodes, "Performance Characteristics of the RVLSI Technology" in G. Saucier and J. Trilhe (eds.), *Proc. IFIP WG 10.5 Workshop on Wafer Scale Integration*, (Grenoble, France), Amsterdam: Elsevier (1986).
7. F.M. Rhodes et al., "A Monolithic Hough Transform Processor Based on Restructurable VLSI," in *IEEE Trans. Pattern Anal. Machine Intell.*, 10, No. 1, 106-110, (1988).
8. S. L. Garverick and E. A. Pierce, "A Single Wafer 16-Point 16 MHz FFT Processor," in *Proc. IEEE Custom Integrated Circuits Conf.*, Rochester, NY, 244-248 (1983), DTIC AD-A147514.
9. J. I. Raffel, "The RVLSI Approach to Wafer-Scale Integration," in *Proc. Workshop on Wafer-Scale Integration*, Southampton Univ., England, (1985).
10. B.-Y. Tsaur, "Assessment of Silicon-on-Insulator Technologies for VLSI," in *Semiconductor-On-Insulator and Thin Film Transistor Technology*, Materials Research Society Symposium Proc. v. 53, Pittsburgh, PA, (1986), DTIC AD-A174015.
11. P. L. Chu, private communication, (1987).
12. P. Denyer and D. Renshaw, *VLSI Signal Processing: A Bit-Serial Approach*, Reading, MA: Addison-Wesley, (1985).
13. R. W. Hendrick, Jr., private communication (1987).
14. H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*, New York: Wiley (1968).
15. K. H. Norsworthy, R. N. Murata, and T. Tibbetts, "Mosaic Signal Processing in Nuclear Environments," Final Report D180-28537-1, Boeing AeroSpace Company (1984).
16. K. H. Norsworthy, R. N. Murata, and B. E. Northon, "Nuclear Induced Noise Mitigation", Technical Report 2-3774-000-181, Boeing AeroSpace Company (1985).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR-860			5. MONITORING ORGANIZATION REPORT NUMBER(S) ESD-TR-89-164		
6a. NAME OF PERFORMING ORGANIZATION Lincoln Laboratory, MIT		6b. OFFICE SYMBOL (If applicable)		7a. NAME OF MONITORING ORGANIZATION Electronic Systems Division	
6c. ADDRESS (City, State, and Zip Code) P.O. Box 73 Lexington, MA 02173-9108			7b. ADDRESS (City, State, and Zip Code) Hanscom AFB, MA 01731		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION HQ AF Systems Command		8b. OFFICE SYMBOL (If applicable) RADC		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-90-C-0002	
8c. ADDRESS (City, State, and Zip Code) Andrews AFB Washington, DC 20334			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 63220C	PROJECT NO. 250	TASK NO. WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Radiation-Hardened Wafer-Scale Integration					
12. PERSONAL AUTHOR(S) Allan H. Anderson, James A Burns, Chenson K. Chen, Gregory J. Dunn, Kenneth H. Konkle, Bor-Yeu Tsaur, Charles E. Woodward, and P.W. Wyatt					
13a. TYPE OF REPORT Technical Report		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Year, Month, Day) 1989, October 25	
15. PAGE COUNT 152					
16. SUPPLEMENTARY NOTATION None					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	CMOS IC processing VLSI fault tolerance focal plane processor gamma circumvention infrared detector nitrided oxides radiation-hardened silicon-on-insulator wafer-scale architecture wafer-scale integration zone melting recrystallization		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>This report describes the logical and physical design of a prototype focal plane processor to be built as a wafer-scale circuit in a radiation-hard CMOS process. Design details and test results are presented for the five circuits which were fabricated in bulk CMOS through the MOSIS fabrication service. The floor plan of the wafer is described. This project was terminated before the wafer-scale circuit could be fabricated. New methods and equipment for fabrication of SOI wafers by zone-melting-recrystallization are described, as well as experience in fabricating CMOS circuits on these wafers. A summary is given of results in related work on fabrication of rad-hard gate dielectrics.</p>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Lt. Col. Hugh L. Southall, USAF			22b. TELEPHONE (Include Area Code) (617) 981-2330		22c. OFFICE SYMBOL ESD/TML